

**UNIVERSIDAD COMPLUTENSE DE MADRID**  
**FACULTAD DE PSICOLOGÍA**



**VALIDEZ DIAGNÓSTICA DE LAS ESCALAS DE LOS  
TRASTORNOS DEPRESIVOS DEL INVENTARIO CLÍNICO  
MULTIAXIAL DE MILLÓN**

**MEMORIA PARA OPTAR AL GRADO DE DOCTOR  
PRESENTADA POR**

**Enrique Chacón Ibarra**

Bajo la dirección del doctor

Jesús Sanz Fernández

**Madrid, 2013**

**UNIVERSIDAD COMPLUTENSE DE MADRID**

**FACULTAD DE PSICOLOGÍA**

**VALIDEZ DIAGNÓSTICA DE LAS  
ESCALAS DE DEPRESIÓN MAYOR Y  
DISTIMIA DEL INVENTARIO CLINICO  
MULTIAXIAL DE MILLON**



**TESIS DOCTORAL**

**Autor: Enrique Chacón Ibarra**

**Director: Jesús Sanz Fernández**



## AGRADECIMIENTOS

Esta tesis doctoral nunca podría haberse escrito sin el apoyo de Jesús Sanz, mi director de tesis, y de María Paz García-Vera, Directora de la Clínica Universitaria de Psicología de la UCM. Gracias a Jesús por acogerme bajo su tutela, guiarme en el diseño y redacción de este trabajo y por hacerse las preguntas básicas que los demás pasamos por alto. Gracias a María Paz por la energía y dedicación que transmite a los que le rodeamos y por confiar en mí en el momento más importante.

Estoy también especialmente agradecido a Ignacio Fernández, que me ha enseñado una buena parte de lo que sé como psicólogo. Eres un gran maestro y amigo.

He tenido la suerte de contar para la realización de esta investigación con un grupo de personas extraordinario, el equipo de psicólogos de la Clínica de Psicología de la Facultad de Psicología de la Universidad Complutense de Madrid. Gracias a María, Paloma, Laura, Noe, Clara, Sara, Natalia, Alejandro, Carmen, Paco, Iván, Miriam, Francisco, Beatriz, Cristina, Raquel, Pablo, Guada, Marta y a los residentes de la clínica que nos precedieron, por vuestra entrega y profesionalidad. Espero que mi futuro esté a vuestro lado.

Gracias a mi madre, mi referencia moral, la persona que siempre está, que me ha empujado a hacer las cosas lo mejor posible, incluida esta tesis doctoral. A mi padre, por ser un modelo en tantas cosas, sobretodo por su capacidad para tomarse la vida con inteligencia, humor y el punto necesario de trascendencia. Gracias a Jose, es una suerte tener a un hermano como tú a mi lado. Gracias por hacerme sentir seguro y por confiar siempre en mí. Gracias a Ágata, eres un ángel. Robert y Carmen, sois la razón por la que trabajo para encontrar un futuro mejor.



También estoy muy agradecido a todos los familiares que me han ayudado en este camino, especialmente a mi abuela y a Fini, que siempre me han dado sabios consejos.

Gracias a Alex, Alba y Rodri, sois más que primos, os siento a mi lado allá donde vaya. Gracias a Miriam y Laura, que me han hecho sentir su apoyo y a Bosco, ese amigo hiperactivo que siempre escucha.

Estoy también agradecido a Ana y a Tere. Siempre recordaré vuestro espíritu de lucha, las ganas de hacer las cosas aun mejor y vuestro sentido de la responsabilidad. Sois una pieza fundamental del funcionamiento de la Clínica y un ejemplo para todos.

Por último, gracias a Regina por su paciencia durante los últimos meses de redacción de esta tesis doctoral, así como por su capacidad para escuchar y apoyar a las personas que forman parte de su vida.

La elaboración de esta tesis doctoral me ha permitido asomarme a la frontera del conocimiento actual de una pequeña área de nuestra ciencia y me ha hecho un mejor psicólogo. Todo esto solo ha sido posible gracias a la Universidad Complutense de Madrid y a la educación pública española, que por encima de todo me ha formado como ciudadano. El sistema educativo es una pieza clave para nuestro futuro como sociedad, mis agradecimientos por tanto a todos los que luchan cada día por defender y mejorar nuestro sistema educativo público y universal.

## Índice de contenidos

<b>Extended Abstract</b> .....	1
<b>PARTE TEÓRICA</b>	
<b>Introducción</b> .....	9
<b>Capítulo 1: Los trastornos depresivos</b>	
1.1. Conceptualización de los trastornos depresivos.....	15
1.1.1. Historia de los estudios sobre la depresión.....	15
1.1.2. Definición y clasificación de los trastornos depresivos.....	22
1.2. Epidemiología e importancia sociosanitaria de los trastornos depresivos...	37
1.2.1. Prevalencia de los trastornos depresivos.....	39
1.2.2. Características epidemiológicas asociadas de los trastornos depresivos.....	54
1.2.3. Importancia sociosanitaria de los trastornos depresivos.....	61
<b>Capítulo 2: El diagnóstico de los trastornos depresivos</b>	
2.1. El papel del diagnóstico en psicología clínica.....	67
2.1.1. Definición de trastorno mental .....	67
2.1.2. Utilidad del diagnóstico de los trastornos mentales.....	70
2.1.3. Fiabilidad del diagnóstico de los trastornos depresivos.....	78
2.1.4. Validez del diagnóstico de los trastornos depresivos.....	79
2.2. El diagnóstico de los trastornos depresivos.....	90
2.2.1 El diagnóstico del trastorno depresivo mayor.....	91
2.2.2 El diagnóstico del trastorno distímico.....	103
2.2.3 Diagnóstico diferencial de los trastornos depresivos.....	105
<b>Capítulo 3: Instrumentos para la evaluación y el diagnóstico de los trastornos depresivos</b>	
3.1. Entrevistas clínicas.....	112

3.2. Las entrevistas diagnósticas estructuradas.....	115
3.3. Cuestionarios, escalas e inventarios para los trastornos depresivos.....	144
3.3.1. Cuestionarios, escalas e inventarios diagnósticos.....	147
3.3.2. Cuestionarios, escalas e inventarios de cribado.....	154
3.3.3. Conclusiones acerca de los cuestionarios, escalas e inventarios para los trastornos depresivos.....	169

## **Capítulo 4. El Inventario Clínico Multiaxial de Millon (MCMI) como instrumento para el diagnóstico de los trastornos depresivos**

4.1. Características generales del Inventario Clínico Multiaxial de Millon (MCMI).....	175
4.2. Las puntuaciones tasa base (TB) o de prevalencia como herramienta diagnóstica.....	180
4.3. Características y versiones del Inventario Clínico Multiaxial de Millon: MCMI-I, MCMI-II y MCMI-III.....	184
4.3.1. Primera versión del Inventario Clínico Multiaxial de Millon (MCMI-I)	184
4.3.2. Inventario Clínico Multiaxial de Millon-II (MCMI-II).....	187
4.3.3. Inventario Clínico Multiaxial de Millon-III (MCMI-III).....	192
4.4. Las escalas para el diagnóstico de los trastornos depresivos del MCMI...	197
4.5. La validez diagnóstica de las escalas para el diagnóstico de los trastornos depresivos del MCMI.....	208
4.6. Problemas de las adaptaciones españolas del MCMI-II y MCMI-III como instrumentos diagnósticos de los trastornos depresivos.....	220

## **PARTE EMPÍRICA**

### **Capítulo 5. Objetivos de la investigación empírica**

5.1. Justificación de la investigación empírica.....	231
5.2. Hipótesis de la investigación empírica.....	233

### **Capítulo 6. Primer estudio: Validez diagnóstica de las escalas de Depresión Mayor y Distimia del MCMI-II**

6.1. Introducción.....	239
6.2. Método.....	243
6.2.1. Participantes.....	243
6.2.2. Instrumentos.....	248
6.2.3. Procedimiento.....	251
6.2.4. Evaluadores.....	252
6.2.5. Diseño y análisis de datos.....	253
6.3. Resultados.....	262
6.4. Discusión.....	320
6.5. Conclusiones.....	338

### **Capítulo 7. Segundo estudio: Validez diagnóstica de las escalas de Depresión Mayor y Distimia del MCMI-III**

7.1. Introducción.....	341
7.2. Método.....	347
7.2.1. Participantes.....	347
7.2.2. Instrumentos.....	350
7.2.3. Procedimiento.....	352
7.2.4. Evaluadores.....	353
7.2.5. Diseño y análisis de datos.....	344
7.4. Resultados.....	356
7.5. Discusión.....	429
7.6. Conclusiones.....	452

### **Capítulo 8. Conclusiones generales..... 457**

### **Referencias bibliográficas..... 467**

### **Anexos..... 505**

## ÍNDICE DE TABLAS

### Capítulo 1: Los trastornos depresivos

<b>Tabla 1.1.</b> Trastornos del estado de ánimo depresivo y trastornos bipolares (APA, 2002).....	26
<b>Tabla 1.2.</b> Síntomas que incluyen los criterios diagnósticos de los trastornos depresivos del DSM-IV y de la CIE-10 (adaptado de Sanz, Izquierdo y García-Vera, en prensa).....	27
<b>Tabla 1.3</b> Prevalencia vital y estimada a la edad de 75 años de los trastornos DSM-IV en la iniciativa WMH de la OMS (Kessler et al., 2007).....	42
<b>Tabla 1.4.</b> Prevalencia Vital y prevalencia a los 12 meses de los trastornos mentales y los trastornos del estado de ánimo en seis países Europeos (Demyttenaere et al., 2004).....	44
<b>Tabla 1.5.</b> Prevalencia de los trastornos mentales y los trastornos depresivos en España según el estudio ESEMeD (Haro et al., 2006a).....	49
<b>Tabla 1.6.</b> Prevalencia vital y a los 12 meses del trastorno depresivo mayor (TDM) y de los trastornos del estado de ánimo (TEA) en los principales estudios epidemiológicos.....	52
<b>Tabla 1.7.</b> Prevalencia a los 12 meses de los trastornos depresivos en hombre y mujeres en estudios europeos de prevalencia.....	55
<b>Tabla 1.8.</b> Prevalencia a los 12 meses del trastorno depresivo mayor según la entrevista CIDI en función de las distintas categorías de gravedad del trastorno y de los diversos países que participaron en la Iniciativa World Mental Health de la Organización Mundial de la Salud.....	57
<b>Capítulo 2: El diagnóstico de los trastornos depresivos</b>	
<b>Tabla 2.1.</b> Especificaciones para el trastorno depresivo mayor según el DSM-IV-TR (APA, 2002).....	102

**Capítulo 3: Instrumentos para la evaluación y el diagnóstico de los trastornos depresivos**

<b>Tabla 3.1.</b> Ventajas y desventajas de las entrevistas estructuradas y semiestructuradas (en Segal, Coolidge, O’Riley y Heinz, 2006, p. 125)	117
<b>Tabla 3.2.</b> Comparación de las características de las entrevistas diagnósticas (modificado de Summerfeldt, Klossterman y Antony, 2011).....	120
<b>Tabla 3.3.</b> Características e índices de fiabilidad (estadístico kappa) para los trastornos depresivos de los estudios sobre la fiabilidad de la entrevista SCID-I (Biometrics Research Department at Columbia University; <a href="http://www.scid4.org">http://www.scid4.org</a> ).....	124
<b>Tabla 3.4.</b> Consistencia del diagnóstico vital DSM-IV entre la entrevista CIDI y la SCID-I en la reevaluación de una muestra del estudio NCS-R (N = 325) (Haro et al., 2006b).....	126
<b>Tabla 3.5.</b> Consistencia del diagnóstico CIDI y SCID-I según el diagnóstico DSM-IV en los últimos 12 meses en la reevaluación clínica de una muestra del estudio ESEMeD y con menos de 60 días de diferencia entre las aplicaciones de las dos entrevistas (n = 143) (Haro et al., 2006b).....	127
<b>Tabla 3.6.</b> Concordancia entre los diagnósticos de trastorno depresivo mayor y distimia de la entrevista MINI-CR (aplicación clínica) y MINI-PR (autoaplicada) con respecto a la SCID-I y la CIDI (Lecrubier et al., 1997, p. 228; Sheehan et al., 1997, p. 235).....	136
<b>Tabla 3.7.</b> Estudios sobre cuestionarios, inventarios y escalas que han sido más citados en el campo de los trastornos depresivos (modificado de Lipsman y Lozano, 2011).....	146
<b>Tabla 3.8.</b> Propiedades diagnósticas del DID en función del sistema de	

corrección seleccionado (Zimmerman, Sheeran y Young, 2004).....	152
<b>Tabla 3.9.</b> Estadísticos de eficacia diagnóstica del DID calculados a partir de dos sistemas de corrección: mediante la suma de la sintomatología o mediante el algoritmo de decisión del DSM-IV (Sheeran y Zimmerman, 2002).....	153
<b>Tabla 3.10.</b> Rangos recomendados para la interpretación del BDI-II (Beck, Steerer y Brown, 1996).....	159
<b>Capítulo 4: El Inventario Clínico Multiaxial de Millon (MCMI) como instrumento para el diagnóstico de los trastornos depresivos</b>	
<b>Tabla 4.1.</b> Características básicas del MCMI-I, MCMI-II y MCMI-III.....	176
<b>Tabla 4.2.</b> Estadísticos de validez diagnóstica de las escalas de Distimia y Depresión Psicótica del MCMI-I del estudio de Wetzler et al. (1989).....	187
<b>Tabla 4.3.</b> Índices de eficacia diagnóstica de las escalas del MCMI-II (punto de corte de TB > 75) para el diagnóstico de los correspondientes síndromes clínicos del Eje I del DSM-III-R (Millon, 2002).....	191
<b>Tabla 4.4.</b> Índices de eficacia diagnóstica de las escalas del MCMI-II (punto de corte de TB ≥ 85) para el diagnóstico de los correspondientes síndromes clínicos del Eje I del DSM-III-R (Millon, 2002).....	191
<b>Tabla 4.5.</b> Cambios que se han producido en las diferentes versiones del MCMI: MCMI-I, MCMI-II y MCMI-III (Rogers, Salekin y Kenneth, 1999)...	191
<b>Tabla 4.6.</b> Índices de eficacia diagnóstica de las escalas del MCMI-III (puntuación de corte de TB ≥ 85) para los correspondientes síndromes clínicos del Eje I del DSM-IV (Millon, Davis y Millon, 2007).....	195
<b>Tabla 4.7.</b> Comparación de los estadísticos medios de validez diagnóstica del MCMI-III publicados en sus dos primeros manuales junto con los estadísticos calculados por Hsu (2002) (medias de todas las escalas)....	197
<b>Tabla 4.8.</b> Número de ítems de las escalas de Depresión Mayor (CC) y Distimia	

(D) del MCMI-II y MCMI-III.....	197
<b>Tabla 4.9.</b> Correspondencia entre los síntomas depresivos clave del diagnóstico de los trastornos depresivos y los ítems de las escalas de trastornos depresivos del MCMI-II y del MCMI-III (adaptado de Sanz, Izquierdo y García-Vera, 2011).....	202
<b>Tabla 4.10.</b> Porcentajes de síntomas depresivos cubiertos por las escalas de trastornos depresivos del MCMI-II y del MCMI-III y porcentajes de ítems de dichas escalas que evalúan síntomas depresivos (adaptado de Sanz, Izquierdo y García-Vera, 2011).....	206
<b>Tabla 4.11.</b> Índices de eficacia diagnóstica de las escalas de Depresión Mayor y Distimia del MCMI-II (puntos de corte de TB $\geq$ 75 y de TB $\geq$ 85) para los correspondientes trastornos depresivos del DSM-III-R (Millon, 2002).....	208
<b>Tabla 4.12.</b> Estadísticos de validez diagnóstica de las escalas de Depresión Mayor y Distimia del MCMI-II para diagnosticar trastorno depresivo mayor en el estudio de Piersna (1991).....	210
<b>Tabla 4.13.</b> Índices de validez diagnóstica de la escalas de Depresión Psicótica o Mayor y de Distimia del MCMI-I y del MCMI-II para diagnosticar un trastorno depresivo en el estudio de Wetzler y Marlowe (1993).....	210
<b>Tabla 4.14.</b> Prevalencia, sensibilidad y poder predictivo positivo (PPP) de las escalas de Distimia y Depresión Mayor del MCMI-III con el punto de corte de TB $\geq$ 85 en el estudio de Millon, 1997).....	211
<b>Tabla 4.15.</b> Estadísticos de validez diagnóstica de las escalas de Depresión Mayor y Distimia del MCMI-III calculados por Rezlaiff (1996) a partir de los datos publicados en el primer manual del MCMI-III (Millon, 1994).....	212
<b>Tabla 4.16.</b> Estadísticos sobre la validez diagnóstica de las escalas de	



Depresión Mayor y Distimia del MCMI-III calculados por Hsu (2002) a partir de los datos de las dos versiones del manual del MCMI-III (Millon, 1994, 1997).....	212
<b>Tabla 4.17.</b> Estadísticos de validez diagnóstica de las escalas de Depresión Mayor y Distimia del MCMI-III en el estudio de Gibeau y Choca (2005)...	213
<b>Tabla 4.18.</b> Estadísticos de validez diagnóstica de las escalas de Depresión Mayor y Distimia del MCMI-III en el estudio de Saulsman (2011).....	215
<b>Tabla 4.19.</b> Estadísticos de validez diagnóstica de la escala de depresión mayor de la adaptación belga del MCMI-III para los diagnósticos asignados con puntuaciones $\geq$ de 1 (trastorno ligero) respecto TB $\geq$ 75 o $\geq$ 5 (presencia de síndrome) respecto TB $\geq$ 85 (Rossi y Sloore, 2005).....	219
<b>Tabla 4.20.</b> Correspondencia en cuanto a la definición de las puntuaciones de TB de la escala de Depresión Mayor (CC) en las muestras normativas de mujeres entre la información que aparece en los baremos de los Anexos del manual de la adaptación española del MCM-II y la información que aparece en las tablas de dicho manual.....	221
<b>Capítulo 5. Objetivos de la investigación empírica</b>	
<b>Tabla 5.1.</b> Características de las dos estudios de la presente investigación....	235
<b>Capítulo 6: Primer estudio: Validez diagnóstica de las escalas de Depresión Mayor y Distimia del MCMI-II</b>	
<b>Tabla 6.1.</b> Índices de validez diagnóstica de las escalas de Depresión Mayor y Distimia del MCMI-II para las puntuaciones de tasa base 75 y 85 en diversos estudios con muestras heterogéneas de pacientes con trastornos psicológicos.....	241
<b>Tabla 6.2.</b> Sexo en la muestra del estudio 1.....	244
<b>Tabla 6.3.</b> Estado civil en la muestra del estudio 1.....	244
<b>Tabla 6.4.</b> Nivel de estudios en la muestra del estudio 1.....	245

<b>Tabla 6.5.</b> Profesión/situación laboral en la muestra del estudio 1.....	245
<b>Tabla 6.6.</b> Demanda asistencial en la muestra del estudio 1.....	246
<b>Tabla 6.7.</b> Diagnósticos según el DSM-IV en la muestra del estudio 1.....	246
<b>Tabla 6.8.</b> Propuesta de interpretación del índice kappa (Altman, 1991).....	256
<b>Tabla 6.9.</b> Áreas bajo la curva ROC en relación con el diagnóstico de trastorno depresivo mayor de las puntuaciones directas y tasa base de las escalas de Depresión Mayor (CC) y Distimia (D) del MCMI-II y de las puntuaciones directas del BDI-II.....	266
<b>Tabla 6.10.</b> Coordenadas de las curvas ROC de las puntuaciones directas y tasa base de la escala de Depresión Mayor (CC) del MCMI-II con respecto al diagnóstico de trastorno depresivo mayor.....	269
<b>Tabla 6.11.</b> Coordenadas de las curvas ROC de las puntuaciones directas y de tasa base de la escala de Distimia del MCMI-II con respecto al diagnóstico de trastorno depresivo mayor.....	272
<b>Tabla 6.12.</b> Coordenadas de las curvas ROC de las puntuaciones del BDI-II con respecto al diagnóstico de trastorno depresivo mayor.....	274
<b>Tabla 6.13.</b> Índices de validez diagnóstica de distintas puntuaciones directas (PD) o tasa base (TB) de la escala de Depresión Mayor (CC) del MCMI-II con respecto al diagnóstico del trastorno depresivo mayor.....	279
<b>Tabla 6.14.</b> Índices de validez diagnóstica de distintas puntuaciones directas (PD) o tasa base (TB) de la escala de Distimia (D) del MCMI-II con respecto al diagnóstico del trastorno depresivo mayor.....	280
<b>Tabla 6.15.</b> Índices de validez diagnóstica de distintas puntuaciones del BDI-II con respecto al diagnóstico de trastorno depresivo mayor.....	281
<b>Tabla 6.16.</b> Áreas bajo la curva ROC en relación con el diagnóstico del conjunto de los trastornos depresivos para las puntuaciones directas y tasa base de las escalas de Depresión Mayor (CC) y Distimia (D) del MCMI-II y de	

las puntuaciones directas del BDI-II.....	285
<b>Tabla 6.17.</b> Coordenadas de las curvas ROC de las puntuaciones directas y de tasa base (TB) de la escala de Depresión Mayor (CC) del MCMI-II con respecto al diagnóstico de cualquier trastorno depresivo.....	288
<b>Tabla 6.18.</b> Coordenadas de las curvas ROC de las puntuaciones directas y de tasa base (TB) de la escala de Distimia (D) del MCMI-II con respecto al diagnóstico de cualquier trastorno depresivo.....	291
<b>Tabla 6.19.</b> Coordenadas de las curvas ROC de las puntuaciones directas del BDI-II con respecto al diagnóstico de cualquier trastorno depresivo.....	293
<b>Tabla 6.20.</b> Índices de validez diagnóstica de distintas puntuaciones directas (PD) o tasa base (TB) de la escala de Depresión Mayor (CC) del MCMI-II con respecto al diagnóstico del conjunto de trastornos depresivos.....	298
<b>Tabla 6.21.</b> Índices de validez diagnóstica de distintas puntuaciones directas (PD) o tasa base (TB) de la escala de Distimia (D) del MCMI-II con respecto al diagnóstico del conjunto de trastornos depresivos.....	299
<b>Tabla 6.22.</b> Índices de validez diagnóstica de distintas puntuaciones del BDI-II con respecto al diagnóstico del conjunto de trastornos depresivos.....	300
<b>Tabla 6.23.</b> Áreas bajo la curva ROC en relación con el diagnóstico de trastorno depresivo mayor (pacientes con el diagnóstico de otros trastornos depresivos excluidos) para las puntuaciones directas y tasa base de las escalas de Depresión Mayor (CC) y Distimia (D) del MCMI-II y para las puntuaciones directas del BDI-II.....	305
<b>Tabla 6.24.</b> Coordenadas de las curvas ROC de las puntuaciones directas y de tasa base (TB) de la escala de Depresión Mayor (CC) del MCMI-II con respecto al diagnóstico de trastorno depresivo mayor (pacientes con el diagnóstico de otros trastornos depresivos excluidos).....	308
<b>Tabla 6.25.</b> Coordenadas de las curvas ROC de las puntuaciones directas y de	

tasa base (TB) de la escala de Distimia (D) del MCMI-II con respecto al diagnóstico de trastorno depresivo mayor (pacientes con el diagnóstico de otros trastornos depresivos excluidos).....	310
<b>Tabla 6.26.</b> Coordinadas de las curvas ROC de las puntuaciones directas del BDI-II con respecto al diagnóstico de trastorno depresivo mayor (pacientes con el diagnóstico de otros trastornos depresivos excluidos)..	313
<b>Tabla 6.27.</b> Índices de validez diagnóstica de distintas puntuaciones directas (PD) o tasa base (TB) de la escala de Depresión Mayor (CC) del MCMI-II con respecto al diagnóstico del trastorno depresivo (pacientes con el diagnóstico de otros trastornos depresivos excluidos).....	317
<b>Tabla 6.28.</b> Índices de validez diagnóstica de distintas puntuaciones directas (PD) o tasa base (TB) de la escala de Distimia (D) del MCMI-II con respecto al diagnóstico del trastorno depresivo mayor (pacientes con el diagnóstico de otros trastornos depresivos excluidos).....	318
<b>Tabla 6.29.</b> Índices de validez diagnóstica de distintas puntuaciones del BDI-II con respecto al diagnóstico de trastorno depresivo mayor (pacientes con el diagnóstico de otros trastornos depresivos excluidos).....	319
<b>Tabla 6.30.</b> Índices de validez diagnóstica de las escalas de Depresión Mayor (CC) y Distimia (D) del MCMI-II para el diagnóstico del trastorno depresivo mayor en muestras heterogéneas de pacientes con trastornos mentales.....	323
<b>Tabla 6.31.</b> Índices de validez diagnóstica de las escalas de Depresión Mayor (CC) y Distimia (D) del MCMI-II para el diagnóstico de cualquier trastorno depresivo en muestras heterogéneas de pacientes con trastornos mentales.....	327

## Capítulo 7. Segundo estudio: Validez diagnóstica de las escalas de Depresión Mayor y Distimia del MCMI-III

<b>Tabla 7.1.</b> Índices de validez diagnóstica de las escalas de Depresión Mayor y Distimia de la versión original del MCMI-III en los estudios de validez realizados para las dos primeras ediciones de su manual.....	342
<b>Tabla 7.2.</b> Índices de validez diagnóstica de las escalas de Depresión Mayor y Distimia del MCMI-III en estudios independientes con muestras heterogéneas de pacientes con trastornos psicológicos.....	344
<b>Tabla 7.3.</b> Sexo en la muestra del estudio 2.....	347
<b>Tabla 7.4.</b> Estado civil en la muestra del estudio 2.....	347
<b>Tabla 7.5.</b> Nivel de estudios en la muestra del estudio 2.....	348
<b>Tabla 7.6.</b> Profesión/ situación laboral en la muestra del estudio 2.....	348
<b>Tabla 7.7.</b> Demanda asistencial en la muestra del estudio 2.....	348
<b>Tabla 7.8.</b> Procedimiento de evaluación diagnóstica del estudio segundo.....	353
<b>Tabla 7.9.</b> Diagnósticos según el DSM-IV en la muestra del estudio segundo....	357
<b>Tabla 7.10.</b> Diagnósticos de trastornos del estado de ánimo según la entrevista diagnóstica estructurada SCID-I en la muestra del estudio segundo.....	358
<b>Tabla 7.11.</b> Análisis de frecuencias de las discrepancias entre el diagnóstico SCID-I para el trastorno depresivo mayor y el diagnóstico de la ficha clínica.....	359
<b>Tabla 7.12.</b> Fiabilidad de las escalas de Depresión Mayor y Distimia del MCMI-III y del BDI-II (alfa de Cronbach).....	360
<b>Tabla 7.13.</b> Correlación de Pearson entre las escalas de Distimia y Depresión	

Mayor del MCMI-III y la puntuación total del BDI-II en el estudio segundo.....	361
<b>Tabla 7.14.</b> Áreas bajo la curva ROC en relación con el diagnóstico de trastorno depresivo mayor para las puntuaciones directas y tasa base de las escalas de Depresión Mayor y Distimia del MCMI-III y de las puntuaciones del BDI-II.....	365
<b>Tabla 7.15.</b> Comparación de las áreas bajo la curva ROC en relación con el diagnóstico de trastorno depresivo mayor para las puntuaciones directas y TB de las escalas de Depresión Mayor (CC) y Distimia (D) del MCMI-III y del BDI-II.....	367
<b>Tabla 7.16.</b> Coordenadas de las curvas ROC de las puntuaciones directas y tasa base de la escala de Depresión Mayor del MCMI-III con respecto al diagnóstico del conjunto de los trastornos depresivos (diagnóstico de la ficha clínica).....	368
<b>Tabla 7.17.</b> Coordenadas de las curvas ROC de las puntuaciones directa y tasa base de la escala de Distimia del MCMI-III con respecto al diagnóstico de trastorno depresivo mayor (diagnóstico de la ficha clínica).....	369
<b>Tabla 7.18.</b> Coordenadas de las curvas ROC de las puntuaciones del BDI-II con respecto al diagnóstico de trastorno depresivo mayor (diagnóstico de la ficha clínica).....	371
<b>Tabla 7.19.</b> Índices de validez diagnóstica de distintas puntuaciones directas (PD) o tasa base (TB) de la escala de Depresión Mayor (CC) del MCMI-III con respecto al diagnóstico de trastorno depresivo mayor (diagnóstico de la ficha clínica).....	376
<b>Tabla 7.20.</b> Índices de validez diagnóstica de distintas puntuaciones directas (PD) o tasa base (TB) de la escala de Distimia (D) del MCMI-III con	

respecto al diagnóstico del trastorno depresivo mayor (diagnóstico de la ficha clínica).....	377
<b>Tabla 7.21.</b> Índices de validez diagnóstica de distintas puntuaciones del BDI-II con respecto al diagnóstico de trastorno depresivo mayor (diagnóstico de la ficha clínica) .....	378
<b>Tabla 7.22.</b> Áreas bajo la curva ROC para el diagnóstico de cualquier trastorno depresivo realizado por los clínicos para las puntuaciones directas y tasa base de las escalas de Depresión Mayor y Distimia del MCMI-III y del BDI-II.....	382
<b>Tabla 7.23.</b> Comparación de las áreas bajo la curva ROC en relación con el conjunto de los trastornos depresivos para las puntuaciones directas y tasa base de las escalas de Depresión Mayor y Distimia del MCMI-III y del BDI-II.....	384
<b>Tabla 7.24.</b> Coordenadas de las curvas ROC de las puntuaciones directas y tasa base de la escala de Depresión Mayor del MCMI-III con respecto al diagnóstico del conjunto de los trastornos depresivos (diagnóstico de la ficha clínica) .....	385
<b>Tabla 7.25.</b> Coordenadas de las curvas ROC de las puntuaciones directas y tasa base de la escala de Distimia del MCMI-III con respecto al diagnóstico del conjunto de los trastornos depresivos (diagnóstico de la ficha clínica) .....	386
<b>Tabla 7.26.</b> Coordenadas de las curvas ROC de las puntuaciones del BDI-II con respecto al diagnóstico del conjunto de los trastornos depresivos (diagnóstico de la ficha clínica) .....	388
<b>Tabla 7.27.</b> Índices de validez diagnóstica de distintas puntuaciones directas (PD) o tasa base (TB) de la escala de Depresión Mayor (CC) del MCMI-III con respecto al diagnóstico del conjunto de los trastornos depresivos	

(diagnóstico de la ficha clínica).....	392
<b>Tabla 7.28.</b> Índices de validez diagnóstica de distintas puntuaciones directas (PD) o tasa base (TB) de la escala de Distimia (D) del MCMI-III con respecto al diagnóstico del conjunto de los trastornos depresivos (diagnóstico de la ficha clínica).....	393
<b>Tabla 7.29.</b> Índices de validez diagnóstica de distintas puntuaciones del BDI-II con respecto al diagnóstico del conjunto de los trastornos depresivos (diagnóstico de la ficha clínica) .....	394
<b>Tabla 7.30.</b> Áreas bajo la curva ROC para el diagnóstico de trastorno depresivo mayor (otros trastornos depresivos excluidos) para las puntuaciones directas y tasa base de las escalas de Depresión Mayor y Distimia del MCMI-III y del BDI-II.....	398
<b>Tabla 7.31.</b> Comparación de las áreas bajo la curva ROC en relación con el diagnóstico de trastorno depresivo mayor (pacientes con otros trastornos depresivos excluidos) para las puntuaciones directas y tasa base de las escalas de Depresión Mayor y Distimia del MCMI-III y para las puntuaciones del BDI-II.....	400
<b>Tabla 7.32.</b> Coordenadas de las curvas ROC de las puntuaciones directas y de tasa base de la escala de Depresión Mayor del MCMI-III con respecto al diagnóstico de trastorno depresivo mayor (pacientes con otros trastornos depresivos excluidos).....	401
<b>Tabla 7.33.</b> Coordenadas de las curvas ROC de las puntuaciones directas y tasa base de la escala de Distimia del MCMI-III con respecto al diagnóstico de trastorno depresivo mayor (pacientes con otros trastornos depresivos excluidos) .....	403
<b>Tabla 7.34.</b> Coordenadas de las curvas ROC de las puntuaciones del BDI-II con respecto al diagnóstico de trastorno depresivo mayor (pacientes con	



otros trastornos depresivos excluidos).....	405
<b>Tabla 7.35.</b> Índices de validez diagnóstica de distintas puntuaciones directas (PD) o tasa base (TB) de la escala de Depresión Mayor (CC) del MCMI-III para el diagnóstico de trastorno depresivo mayor (excluidos pacientes con otros trastornos depresivos).....	408
<b>Tabla 7.36.</b> Índices de validez diagnóstica de distintas puntuaciones directas (PD) o tasa base (TB) de la escala de Distimia (D) del MCMI-III para el diagnóstico del trastorno depresivo mayor (excluidos pacientes con otros trastornos depresivos).....	409
<b>Tabla 7.37.</b> Índices de validez diagnóstica de distintas puntuaciones del BDI-II con respecto al diagnóstico de trastorno depresivo mayor (excluidos pacientes con el diagnóstico de otros trastornos depresivos) .....	410
<b>Tabla 7.38.</b> Áreas bajo la curva ROC en relación con el diagnóstico de trastorno depresivo mayor (diagnóstico SCID-I) para las puntuaciones directas y tasa base de las escalas de Depresión Mayor y Distimia del MCMI-III y para las puntuaciones del BDI-II.....	415
<b>Tabla 7.39.</b> Comparación de las áreas bajo la curva ROC en relación con el diagnóstico de trastorno depresivo mayor de la entrevista SCID-I para las puntuaciones directas y tasa base de las escalas de Depresión Mayor (CC) y Distimia (D) del MCMI-III y para las puntuaciones directas del BDI-II.....	417
<b>Tabla 7.40.</b> Coordenadas de las curvas ROC de las puntuaciones directas y tasa base de la escala de Depresión Mayor del MCMI-III con respecto al diagnóstico de trastorno depresivo mayor de la SCID-I.....	418
<b>Tabla 7.41.</b> Coordenadas de las curvas ROC de las puntuaciones directas y tasa base de la escala de Distimia del MCMI-III con respecto al diagnóstico de trastorno depresivo mayor de la SCID-I.....	420

<b>Tabla 7.42.</b> Coordenadas de las curvas ROC de las puntuaciones del BDI-II con respecto al diagnóstico de trastorno depresivo mayor de la SCID-I...	422
<b>Tabla 7.43.</b> Índices de validez diagnóstica de distintas puntuaciones directas (PD) o tasa base (TB) de la escala de Depresión Mayor (CC) del MCMI-III con respecto al diagnóstico de trastorno depresivo mayor (diagnóstico SCID-I).....	426
<b>Tabla 7.44.</b> Índices de validez diagnóstica de distintas puntuaciones directas (PD) o tasa base (TB) de la escala de Distimia (D) del MCMI-III con respecto al diagnóstico de trastorno depresivo mayor (diagnóstico SCID-I)	427
<b>Tabla 7.45.</b> Índices de validez diagnóstica de distintas puntuaciones del BDI-II con respecto al diagnóstico de trastorno depresivo mayor (diagnóstico SCID-I) .....	428
<b>Tabla 7.46.</b> Índices de validez para el diagnóstico del trastorno depresivo mayor obtenidos en este estudio para las escalas de Depresión Mayor y Distimia del MCMI-III.....	433
<b>Tabla 7.47.</b> Índices de validez diagnóstica para el diagnóstico de cualquier trastorno depresivo de las escalas de Depresión Mayor y Distimia del MCMI-III en muestras de pacientes con trastornos psicológicos.....	437

## ÍNDICE DE FIGURAS

### Capítulo 4: El Inventario Clínico Multiaxial de Millon (MCMI) como instrumento para el diagnóstico de los trastornos depresivos

- Figura 4.1.** Búsqueda en los campos “resumen”, “descriptores” y “pruebas y medidas” de PsycINFO de “MCMI” o “Millon Clinical Multiaxial Inventory” 177
- Figura 4.2.** Búsqueda en los campos “resumen”, “descriptores” y “pruebas y medidas” de PsycINFO de “MMPI” o “Minnesota Multiphasic Personality Inventory”, “MCMI” o “Millon Clinical Multiaxial Inventory”, “Rorschach” y “SCL-90” o “Symptom Checklist-90” ..... 178

### Capítulo 6: Primer estudio: Validez diagnóstica de las escalas de Depresión Mayor y Distimia del MCMI-II

- Figura 6.1.** Plano sobre el que se sitúa el gráfico del área bajo la curva ROC (denominada curva COR en el paquete estadístico SPSS)..... 259
- Figura 6.2.** Curva ROC de las puntuaciones directas de la escala de Depresión Mayor (CC) del MCMI-II con respecto al diagnóstico de trastorno depresivo mayor..... 263
- Figura 6.3.** Curva ROC de las puntuaciones tasa base (TB) de la escala de Depresión Mayor (CC) del MCMI-II con respecto al diagnóstico trastorno depresivo mayor..... 263
- Figura 6.4.** Curva ROC de las puntuaciones directas de la escala de Distimia (D) del MCMI-II con respecto al diagnóstico de trastorno depresivo mayor..... 264
- Figura 6.5.** Curva ROC de las puntuaciones tasa base (TB) de la escala de Distimia (D) del MCMI-II con respecto al diagnóstico de trastorno depresivo mayor..... 264

<b>Figura 6.6.</b> Curva ROC de las puntuaciones del BDI-II con respecto al diagnóstico de trastorno depresivo mayor.....	265
<b>Figura 6.7.</b> Curvas ROC de las puntuaciones directas de las escalas de Distimia (D) y Depresión Mayor (CC) del MCMI-II y del BDI-II con respecto al diagnóstico de trastorno depresivo mayor.....	268
<b>Figura 6.8.</b> Curvas ROC de las puntuaciones tasa base de las escalas de Distimia (D) y Depresión Mayor (CC) del MCMI-II y de las puntuaciones directas del BDI-II con respecto al diagnóstico de trastorno depresivo mayor.....	268
<b>Figura 6.9.</b> Curva ROC de las puntuaciones directas de la escala de Depresión Mayor (CC) del MCMI-II con respecto al diagnóstico de cualquier trastorno depresivo.....	282
<b>Figura 6.10.</b> Curva ROC de las puntuaciones tasa base de la escala de Depresión Mayor (CC) del MCMI-II con respecto al diagnóstico de cualquier trastorno depresivo.....	283
<b>Figura 6.11.</b> Curva ROC de las puntuaciones directas de la escala de Distimia (D) del MCMI-II con respecto al diagnóstico de cualquier trastorno depresivo.....	283
<b>Figura 6.12.</b> Curva ROC de las puntuaciones tasa base de la escala de Distimia (D) del MCMI-II con respecto al diagnóstico de cualquier trastorno depresivo.....	284
<b>Figura 6.13.</b> Curva ROC de las puntuaciones directas del BDI-II con respecto al diagnóstico de cualquier trastorno depresivo.....	284
<b>Figura 6.14.</b> Curvas ROC de las puntuaciones directas de las escalas de Distimia (D) y de Depresión Mayor (CC) del MCMI-II y de las puntuaciones directas del BDI-II con respecto al diagnóstico del conjunto de trastornos depresivos.....	287

<b>Figura 6.15.</b> Curvas ROC de las puntuaciones tasa base (TB) de las escalas de Distimia (D) y de Depresión Mayor (CC) del MCMI-II y de las puntuaciones directas del BDI-II con respecto al diagnóstico del conjunto de trastornos depresivos.....	287
<b>Figura 6.16.</b> Curva ROC de las puntuaciones directas de la escala de Depresión Mayor (CC) del MCMI-II con respecto al diagnóstico de trastorno depresivo mayor (pacientes con el diagnóstico de otros trastornos depresivos excluidos) .....	301
<b>Figura 6.17.</b> Curva ROC de las puntuaciones tasa base (TB) de la escala de Depresión Mayor (CC) del MCMI-II con respecto al diagnóstico trastorno depresivo mayor (pacientes con el diagnóstico de otros trastornos depresivos excluidos) .....	302
<b>Figura 6.18.</b> Curva ROC de las puntuaciones directas de la escala de Distimia (D) del MCMI-II con respecto al diagnóstico de trastorno depresivo mayor (pacientes con el diagnóstico de otros trastornos depresivos excluidos) .....	302
<b>Figura 6.19.</b> Curva ROC de las puntuaciones tasa base (TB) de la escala de Distimia (D) del MCMI-II con respecto al diagnóstico de trastorno depresivo mayor (pacientes con el diagnóstico de otros trastornos depresivos excluidos) .....	303
<b>Figura 6.20.</b> Curva ROC de las puntuaciones del BDI-II con respecto al diagnóstico de trastorno depresivo mayor (pacientes con el diagnóstico de otros trastornos depresivos excluidos).....	303
<b>Figura 6.21.</b> Curvas ROC de las puntuaciones directas de la escala de Distimia y Depresión Mayor del MCMI-II con respecto al diagnóstico de Trastorno Depresivo Mayor (pacientes con el diagnóstico de otros trastornos depresivos excluidos) .....	306

<b>Figura 6.22.</b> Curvas ROC de las puntuaciones tasa base de las escalas de Distimia y Depresión Mayor del MCMI-II con respecto al diagnóstico de trastorno depresivo mayor (pacientes con el diagnóstico de otros trastornos depresivos excluidos) .....	307
<b>Capítulo 7. Segundo estudio: Validez diagnóstica de las escalas de Depresión Mayor y Distimia del MCMI-III</b>	
<b>Figura 7.1.</b> Curva ROC de las puntuaciones directas de la escala CC del MCMI-III para el diagnóstico de trastorno depresivo mayor (diagnóstico de la ficha clínica) .....	362
<b>Figura 7.2.</b> Curva ROC de las puntuaciones TB de la escala CC del MCMI-III con respecto al diagnóstico de trastorno depresivo mayor (diagnóstico de la ficha clínica) .....	363
<b>Figura 7.3.</b> Curva ROC de las puntuaciones directas de la escala de D del MCMI-III con respecto al diagnóstico de trastorno depresivo mayor (diagnóstico de la ficha clínica) .....	363
<b>Figura 7.4.</b> Curva ROC de las puntuaciones tasa base de la escala D del MCMI-III con respecto al diagnóstico de trastorno depresivo mayor (diagnóstico de la ficha clínica) .....	364
<b>Figura 7.5.</b> Curva ROC de las puntuaciones del BDI-III con respecto al diagnóstico de trastorno depresivo mayor (diagnóstico de la ficha clínica).....	364
<b>Figura 7.6.</b> Curvas ROC de las puntuaciones directas y TB de las escalas de Distimia y Depresión Mayor del MCMI-III y del BDI-II con respecto al diagnóstico de trastorno depresivo mayor (diagnóstico de la ficha clínica) .....	366
<b>Figura 7.7.</b> Curva ROC de las puntuaciones directas de la escala de CC del MCMI-III con respecto al diagnóstico de cualquier trastorno depresivo	

(diagnóstico de la ficha clínica).....	379
<b>Figura 7.8.</b> Curva ROC de las puntuaciones tasa base de la escala de CC del MCMI-III con respecto al diagnóstico cualquier trastorno depresivo (diagnóstico de la ficha clínica).....	380
<b>Figura 7.9.</b> Curva ROC de las puntuaciones directas de la escala de D del MCMI-III con respecto al diagnóstico de cualquier trastorno depresivo (diagnóstico de la ficha clínica).....	380
<b>Figura 7.10.</b> Curva ROC de las puntuaciones tasa base de la escala de D del MCMI-III con respecto al diagnóstico de cualquier trastorno depresivo (diagnóstico de la ficha clínica).....	381
<b>Figura 7.11.</b> Curva ROC de las puntuaciones del BDI-II con respecto al diagnóstico de cualquier trastorno depresivo (diagnóstico de la ficha clínica).....	382
<b>Figura 7.12.</b> Curvas ROC de las puntuaciones directas y TB de la escala de Distimia y Depresión Mayor del MCMI-III con respecto cualquier trastorno depresivo (diagnóstico de la ficha clínica).....	383
<b>Figura 7.13.</b> Curva ROC de las puntuaciones directas de la escala de Depresión Mayor del MCMI-III con respecto al diagnóstico de trastorno depresivo mayor (pacientes con otros trastornos depresivos excluidos; diagnóstico de la ficha clínica).....	395
<b>Figura 7.14.</b> Curva ROC de las puntuaciones TB de la escala de Depresión Mayor del MCMI-III con respecto al diagnóstico trastorno depresivo mayor (pacientes con otros trastornos depresivos excluidos; diagnóstico de la ficha clínica).....	396
<b>Figura 7.15.</b> Curva ROC de las puntuaciones directas de la escala de Distimia del MCMI-III con respecto al diagnóstico de trastorno depresivo mayor (pacientes con otros trastornos depresivos excluidos; diagnóstico	

de la ficha clínica).....	396
<b>Figura 7.16.</b> Curva ROC de las puntuaciones TB de la escala de Distimia del MCMI-III con respecto al diagnóstico de trastorno depresivo mayor (pacientes con otros trastornos depresivos excluidos; diagnóstico de la ficha clínica).....	397
<b>Figura 7.17.</b> Curva ROC de las puntuaciones del BDI-II con respecto al diagnóstico de trastorno depresivo mayor (pacientes con otros trastornos depresivos excluidos; diagnóstico de la ficha clínica).....	397
<b>Figura 7.18.</b> Comparación de las curvas ROC de las puntuaciones directas y TB de la escalas de Distimia y Depresión Mayor del MCMI-III y de las puntuaciones del BDI-II con respecto al diagnóstico de trastorno depresivo mayor (pacientes con otros trastornos depresivos excluidos)..	399
<b>Figura 7.19.</b> Curva ROC de las puntuaciones directas de la escala de CC del MCMI-III con respecto al diagnóstico de trastorno depresivo mayor (diagnóstico SCID-I) .....	412
<b>Figura 7.20.</b> Curva ROC de las puntuaciones tasa base de la escala de CC del MCMI-III con respecto al diagnóstico de trastorno depresivo mayor (diagnóstico SCID-I) .....	412
<b>Figura 7.21.</b> Curva ROC de las puntuaciones directas de la escala de D del MCMI-III con respecto al diagnóstico de trastorno depresivo mayor (diagnóstico SCID-I) .....	413
<b>Figura 7.22.</b> Curva ROC de las puntuaciones tasa base de la escala de D del MCMI-III con respecto al diagnóstico de Trastorno depresivo mayor (diagnóstico SCID-I) .....	413
<b>Figura 7.23.</b> Curva ROC de las puntuaciones del BDI-II con respecto al diagnóstico de trastorno depresivo mayor (diagnóstico SCID-I).....	414
<b>Figura 7.24.</b> Curvas ROC del BDI-II, las puntuaciones directas y TB de la	



escala de Distimia y Depresión Mayor del MCMI-III con respecto al diagnóstico de trastorno depresivo mayor (Diagnóstico SCID-I) .....	416
------------------------------------------------------------------------------------------------------------------------------------------	-----

## **Extended Abstract**

### **Diagnostic validity of the Millon Clinical Multiaxial Inventory Major Depression and Dysthymia scales**

#### **INTRODUCTION**

Depressive disorders are common conditions widely dispersed among the population and are usually associated with a high degree of disability (Ebmeier, Donaghey & Steele, 2006). Nowadays, there are no biological markers that allow clinicians to make a reliable and valid diagnosis of depressive disorders. Thus, the "gold standard" to diagnose these conditions is the diagnosis obtained with structured diagnostic interviews administered by trained clinicians.

In the last 30 years, diagnostic activity has dramatically risen and diagnostic systems such as DSM-IV are widely used. Clinicians have numerous questionnaires, inventories and scales to assess depressive symptoms. In fact, some of them have been specifically created for the diagnosis of depressive disorders. As a result, we should be rigorous in choosing the correct instruments to make a diagnosis.

The Millon Clinical Multiaxial Inventory (MCMI; Millon, 1983) is one of the most used instruments in Europe (Evers, 2012) and Spain (Muñiz & Fernández-Hermida, 2010). There are two frequently used versions: the MCMI-II (Millon, 1987) and the MCMI-III (Millon, 1994; Millon, Davis & Millon, 1997; Millon, Millon, Davis & Grossman, 2009). One of the main aims of the MCMI is the detection and diagnosis of mental disorders, including major depressive disorder and dysthymic disorder. At the moment, there are no published studies on the diagnostic validity of the Spanish

adaptation of the MCMI-II (Millon, 1998, 1999, 2002, Spanish adaptation by A. Ávila-Espada) or the diagnostic validity of the Spanish adaptation of the MCMI-III (Millon, Millon & Davis, 2007, Spanish adaptation by V. Cardenal and M. P. Sánchez). In addition, studies that have evaluated the diagnostic efficacy of the original versions of the MCMI-II and MCMI-III have revealed methodological problems (Millon, 1994; Millon et al., 1997). Moreover, the diagnostic validity of these inventories has been the subject of great controversy (Hsu, 2002; Retzlaff, 1996; Grove & Vrieze, 2009). Therefore, the general aim of this research was to examine the diagnostic efficacy of the MCMI-II/MCMI-III Major Depression and Dysthymia scales in diagnosing DSM-IV depressive disorders.

## **METHODOLOGY**

The specific aims of this research were firstly to examine the diagnostic validity of the base rate (BR) score of 75 and the BR score of 85 on the Major Depression (CC) and Dysthymia (D) scales (the standard cut-off scores in the MCMI-II and MCMI-III) when they are used to diagnosis DSM-IV major depressive disorder or when they are used to diagnosis any DSM-IV depressive disorder (e.g., major depressive disorder, dysthymia, depressive disorder not otherwise specified, adjustment disorder with depressed mood, bereavement). Secondly, this research examined whether there were other scores on the MCMI-II/MCMI-III Major Depression and Dysthymia scales with better diagnostic efficacy than the standard cut-off scores. Finally, we also examined whether diagnostic efficacy indices for the MCMI-II/MCMI-III Major Depression and Dysthymia scales were better than those for other instruments widely used to assess depression, in particular better than those for the Beck Depression Inventory-II (BDI-II; Beck et al., 1996). Thus, two independent psychometric studies were conducted with the MCMI-II and MCMI-III.

The first study examined the diagnostic validity of the MCMI-II and was an instrumental retrospective study, conducted with a sample of 181 outpatients with a variety of mental disorders. Psychologists generally knew both the MCMI-II and BDI-II scores of the patients before making the diagnosis. In consequence, it was not possible to exclude biases or cross-contamination between the scores of these inventories and the diagnoses made by psychologists. The instruments used in the first study were the third edition of the Spanish adaptation of the MCMI-II (Millon, 2002, Spanish adaptation by A. Ávila-Espada), the Spanish adaptation of the BDI-II (Sanz et al., 2005; Sanz and García-Vera, 2013; Sanz, Navarro & Vazquez, 2003; Sanz, Perdigón & Vázquez, 2003), and the Checklist for Depressive Episode (LED; Sanz et al., 2005).

The second study was an instrumental prospective study designed to examine the diagnostic validity of the MCMI-III. This study was carried out with 203 outpatients with a variety of mental disorders who filled out the MCMI-III and BDI-II when they were still undiagnosed. All patients underwent a structured diagnostic interview (the mood disorder module of the Structured Clinical Interview for DSM-IV Axis I Disorders, Clinician Version, or SCID-I; First, Spitzer, Gibbon & Williams, 1999), as well as other assessment procedures. After 3-4 psychological assessment sessions, a diagnosis was made by psychologists and recorded in diagnostic sheets. This clinical diagnosis was used as a reference diagnostic criterion, but also the diagnosis provided by the SCID-I was used as a reference diagnostic criterion. The instruments used in the second study were the Spanish adaptation of the MCMI-III (Millon, Davis & Millon, 2007, Spanish adaptation by V. Cardenal & M. P. Sánchez), the Spanish adaptation of the BDI-II, and the Spanish version of the SCID-I.

A variety of diagnostic efficacy statistics were computed for each diagnosis (major depressive disorder; any depressive disorder), each reference diagnostic criterion (clinical diagnosis; SCID-I diagnosis), each type of MCMI scores (BR scores; raw scores), and each instrument (the MCMI-II/MCMI-III Major Depressive scale; the MCMI-II/MCMI-III Dysthymia scale; the BDI-II): sensitivity, specificity, Youden index, kappa coefficient (diagnostic agreement), positive predictive value, negative predictive value, and area under the ROC (receiver operating characteristic) curve (or AUC). In addition, *z* statistics were computed to test differences between the AUCs of two instruments.

## RESULTS

According to the results of both studies, the diagnostic efficacy of the BR scores of 75 and 85 of the MCMI-II/MCMI-III Major Depression and Dysthymia scales is questionable. The majority of their diagnostic efficacy statistics were below the standards for a diagnostic instrument, both when diagnosing major depressive disorder and when diagnosing any depressive disorder. In particular, the BR scores of 75 and 85 showed sensitivities below the standard of 70%, kappa coefficients below the standard of 40, and Youden indices below the standard of 40 (this last value reflects an appropriate balance between sensitivity and specificity). In general, only specificities (> 70%) and negative predictive values (> 50%) supported the diagnostic validity of the standard cut-off scores of the MCMI-II/MCMI-III CC and D scales (BR 75 and BR 85) in diagnosing major depressive disorder or any depressive disorder. Therefore, it should be noted that by using BR scores of 75 or 85, a high percentage of patients who really suffer from depressive disorders would be left undetected. However, those cut-off scores could be useful to exclude the presence of a depressive disorder.

These results are poorer than those reported in the original MCMI-II/MCMI-III manuals (Millon et al., 1997, 2002), but they are consistent with those reported by previous independent studies (Gibeau & Choca, 2005; Piersna, 1991; Rossi & Sloore, 2005; Saulsman, 2011; Wetzler & Marlowe, 1993). These last studies also showed that the diagnostic efficacy statistics for the standard cut-off scores (BR 75 and BR 85) of the MCMI-II/MCMI-III CC and D scales were, in general, below the standard values for a diagnostic tool.

However, when diagnosing a major depressive disorder, most of the AUCs of the MCMI-II/MCMI-III CC and D scales were statistically significant and higher than the standard for a diagnostic tool ( $> 0.70$ ; Swets, 1988), even than the most stringent standard ( $> 0.75$ ; Fan et al., 2006). Therefore, it is not surprising that other cut-off BR or raw scores different from the BR scores of 75 and 85 showed similar or slightly higher diagnostic efficacy indices than did the BR scores of 75 and 85. This result suggests that the methodological strategy of using BR scores to obtain the cut-off scores of a diagnostic instrument does not always assure that the best cut-off scores are obtained. Furthermore, it does not always assure cut-off scores with appropriate indices of diagnostic validity. Therefore, it is advisable to use other methodological strategies (e.g., the ROC analysis) that examine the diagnostic performance of a wide range of scores in order to select those scores with the best indices of diagnostic validity.

For the MCMI-II, although it was possible to find alternative cut-off scores with diagnostic efficacy indices slightly higher than those of the BR scores of 75 and 85, in general, those alternative cut-off scores did not exceed the standards for a diagnostic tool.

However, for the MCMI-III, especially with the SCID-I diagnosis was used as the reference diagnostic criterion for major depression disorder, it was possible to

find several alternative cut-off points that not only showed higher diagnostic efficacy indices than did the BR scores of 75 and 85, but also they showed appropriate indices of diagnostic validity: kappa > 0.40, Youden index > 40, sensitivity > 70%, specificity > 70%, positive and negative predictive values > 50%. These results are poorer than those reported in the second edition of the MCMI-III manual (Millon et al., 1997), but they are better than those found in previous studies (Gibeau & Choca, 2005; Rossi & Sloore, 2005), even when the reference diagnostic criterion was also obtained by means of a structured diagnostic interview (Saulsman, 2011).

On the other hand, the results of both studies also suggest that the diagnostic performance of the MCMI-II/MCMI-III Major Depression and Dysthymia scales is not better than that of the BDI-II, but very similar. There were no significant differences between the AUCs of the MCMI-II/MCMI-III scales and those of the BDI-II. In addition, the diagnostic validity indices of the MCMI-II/MCMI-III scales for both the diagnosis of major depressive disorder and the diagnosis of any depressive disorder were also very similar to those of the BDI-II. In fact, with the SCID-I diagnosis was used as the reference diagnostic criterion for major depression disorder, it was possible to find several BDI-II cut-off points that showed appropriate indices of diagnostic validity: kappa > 0.40, Youden index > 40, sensitivity > 70%, specificity > 70%, positive and negative predictive values > 50%.

## **CONCLUSIONS**

This research had some limitations. Both in the MCMI-II study and the MCMI-III study, the inter-rater reliability of the clinical diagnosis was not examined. In addition, clinicians in the MCMI-II study usually knew both the MCMI-II and BDI-II scores of the patients before making the diagnosis and the diagnoses were not based on a structured diagnostic interview. However, in spite of those limitations, the results of the present research indicate that the standard cut-off scores (BR 75

and BR 85) of the Major Depression and Dysthymia scales of the Spanish adaptations of the MCMI-II/MCMI-III have questionable diagnostic validity. Using the standard cut-off scores, the MCMI-II/MCMI-III depressive disorder scales have particular difficulties in detecting patients suffering from major depressive disorder or patients suffering from any depressive disorder, but they could be useful to exclude the presence of a depressive disorder.

On the other hand, the results of this research also point out that, at least the Major Depression and Dysthymia scales of the Spanish adaptation of the MCMI-III could be useful to diagnose a major depressive disorder or to diagnose any depressive disorder insofar as appropriate cut-off points may be set and used. For example, the MCMI-III CC and D scales could be useful to diagnose depressive disorders if the valid cut-off points found in the present research were replicated in future studies.

Clearly, more thorough research is needed for a better understanding of the diagnostic performance of the MCMI-II and MCMI-III. Therefore, it is currently not recommended to use the MCMI-II/MCMI-III Major Depression or Dysthymia scales as principal instruments for the diagnosis of depressive disorders. Psychopathological diagnosis should be based on the application of a structured diagnostic interview, especially when that diagnosis is the main purpose of assessment.





# Parte Teórica



## INTRODUCCIÓN

Los trastornos depresivos se caracterizan por la presencia de bajo estado de ánimo persistente o disminución del placer o interés, asociado a un conjunto complejo de síntomas somáticos, cognitivos, conductuales y motivacionales (p. ej, apatía, pérdida de apetito, cambios del sueño, enlentecimiento psicomotor, sentimientos de infravaloración o culpa, etc.). Se considera que esta sintomatología configura un trastorno cuando por su frecuencia, duración o intensidad, provoca un malestar significativo en la persona o deteriora su rendimiento laboral habitual, sus relaciones sociales u otras áreas importantes de su actividad. Determinar si el descenso del estado de ánimo o del placer o interés y los síntomas que lo acompañan configuran en una persona un trastorno mental es una tarea compleja (Andrews et al., 2007).

No se han encontrado hasta ahora marcadores biológicos o criterios causales inequívocos que permitan llegar a un diagnóstico totalmente fiable y válido de los trastornos depresivos, de manera que el “patrón de oro” para su detección, al igual que ocurre con la mayoría de los trastornos mentales, es la aplicación por un clínico con experiencia de una entrevista estructurada que evalúe, fundamentalmente, la presencia, curso, gravedad y disfuncionalidad de los síntomas que se consideran parte del constructo de trastorno depresivo. Se dispone además de un amplio abanico de cuestionarios, inventarios y escalas que evalúan la sintomatología depresiva, los cuales pueden ayudar a la hora de realizar un diagnóstico de trastorno depresivo. Es más, algunos de estos cuestionarios, inventarios y escalas han sido elaborados específicamente para estos fines diagnósticos.

Uno de los instrumentos más utilizados tanto en la práctica como en la investigación en psicología clínica es el Inventario Clínico Multiaxial de Millon

(MCMI; Millon, 1983), especialmente sus dos últimas versiones, el MCMI-II (Millon, 1987) y el MCMI-III (Millon, 1994; Millon, Davis y Millon, 1997; Millon, Millon, Davis y Grossman, 2009). Se trata de un test complejo, que en sus sucesivas versiones se ha modificado para acercarse a los criterios diagnósticos del DSM-IV (APA, 2002). Precisamente uno de los principales objetivos del MCMI es la detección y diagnóstico de los trastornos mentales, entre ellos el trastorno depresivo mayor y el trastorno distímico. Los autores de este instrumento han llamado continuamente a la investigación sobre sus características (Craig, 2005b), pero no se han publicado hasta ahora estudios sobre la validez diagnóstica de las adaptaciones españolas del MCMI-II (Millon, 1998, 1999, 2002, adaptación española de A. Ávila-Espada) o del MCMI-III (Millon, Davis y Millon, 2007, adaptación española de V. Cardenal y M. P. Sánchez), por lo que el objetivo principal de la presente tesis doctoral fue comprobar las cualidades psicométricas de sus escalas de Depresión Mayor y Distimia para la detección de los trastornos depresivos.

La correcta detección de los trastornos depresivos es fundamental en el campo de la salud mental. Los estudios epidemiológicos de las últimas décadas han confirmado que estos trastornos tienen una alta prevalencia, son muy incapacitantes y producen un alto coste social (Ebmeier, Donaghey y Steele, 2006). Debido a esa alta prevalencia, al aumento de las tasas de depresión en todas las edades y al descenso en la edad de su inicio, se ha señalado que estamos en “la era de la depresión” (Klerman y Weissman, 1989). De hecho, “la depresión es el diagnóstico que se emplea con más frecuencia y el que potencialmente tiene más peligro de mortalidad debido a las conductas suicidas” (Vázquez y Sanz, 2008, p. 4).

# **Capítulo 1.**

## **LOS TRASTORNOS DEPRESIVOS**



**“La Torre de Babel nunca produjo tanta confusión de lenguas como la variedad de síntomas que produce el caos de los melancólicos”**

**Robert Burton, 1621/2008, p. 228.**

### **1.1. Conceptualización de los trastornos depresivos**

Definir los trastornos depresivos es una tarea compleja que no en vano, y como se verá más adelante, ha sido objeto de importantes discusiones a lo largo de la historia de la psiquiatría y de la psicología clínica. Tanto en la práctica clínica como incluso a nivel teórico o de investigación, el concepto de depresión aparece definido de manera poco clara y aludiendo a fenómenos que, aunque relacionados, son diferentes. En ocasiones la depresión se entiende como un estado de ánimo, otras como un síntoma, a veces aparece como un síndrome y en otras ocasiones directamente como un trastorno (García-Vera y Sanz, 2012; Vázquez y Sanz, 1991, 2008). Respecto a lo que parece que hay consenso es que se trata de un concepto complejo en cuanto a sus acepciones. No en vano, el constructo depresión entendido como trastorno, es decir, como un conjunto de síntomas relacionados que aparecen y desaparecen conjuntamente y para el cual se han especificado ciertos parámetros de duración, gravedad, curso o historia natural y, además, se han descartado algunas posibles causas, ha ido sufriendo transformaciones a lo largo del tiempo y las diferentes épocas.

#### **1.1.1. Historia de los estudios sobre la depresión**

Existen en la literatura científica diversos análisis históricos sobre el estudio de los trastornos del estado de ánimo (Berrios, 2008; Conti, 2007; Hirshbein, 2009; Jackson 1989; Millon y Simonsen, 2010). En este epígrafe se presenta información que puede ayudar a entender el estado actual de conocimientos sobre los trastornos depresivos.



Se puede encontrar en la Grecia clásica las primeras referencias explícitas basadas en la observación sobre el estudio de los cambios en el estado de ánimo (Millon y Simonsen, 2010). Aunque existen elementos artísticos anteriores sobre la tristeza y la melancolía, el primer acercamiento científico fue realizado por la escuela hipocrática. Su principal representante, Hipócrates de Cos (Siglo IV a.C.) desarrolló la teoría de los humores según la cual en el organismo fluyen cuatro elementos esenciales: la sangre, la bilis negra, la bilis amarilla y la flema. Se consideraba que las variaciones en estos elementos, debido a razones internas o externas, eran las causas de las enfermedades. En concreto, la melancolía se produciría cuando hay un exceso de bilis negra en el cuerpo del paciente, lo que provocaría una serie de síntomas como abatimiento, tristeza, ansiedad, dificultades para dormir, aislamiento, irritabilidad... Con el paso de los siglos, existió una confusión con este término (melancolía, *melaina chole*, bilis negra), puesto que con frecuencia su uso hacía referencia a distintos estados anormales: tristeza con desesperanza, tristeza con angustia, hipocondría o delirios. Durante siglos la melancolía estuvo asociada al temperamento sombrío, a pesar de que los escritos Aristotélicos (384-322 a. C.) le otorgó unas capacidades positivas (idea que se recuperaría más tarde durante el romanticismo) (Jackson, 1989). Así, para Platón (siglo IV a.C.), el dolor y la tristeza era considerados como una enfermedad en la medida en la que disminuía la capacidad de razonamiento (Jackson, 1989).

Los conocimientos del mundo helénico se trasladaron a Roma, donde las distintas escuelas debatían sobre los orígenes de la enfermedad, los cuales se situaban en el esófago, en la cabeza o en el estómago según cada escuela. Algunos autores romanos destacaban la importancia de tratar la enfermedad desde el inicio del problema y por ello subrayaban la necesidad de la detección temprana. Además se postulaban causas endógenas (de nacimiento o adquiridas) y causas

ambientales. Galeno (siglo II d.C.) destacó la importancia del miedo y el humor depresivo duradero, como elementos que llevaban a los pacientes a la melancolía. Los factores que influían según este autor en la formación de estado de ánimo eran ambientales más que psicológicos: el aire, el ejercicio, el descanso, el sueño y la vigilia, la comida y la bebida, la excreción de elementos superfluos y las pasiones o perturbaciones del espíritu (Millon y Simonsen, 2010).

En el mundo cristiano, los estudiosos de la medicina formaban parte de la iglesia, de manera que una parte de las explicaciones sobre las enfermedades mentales tenían relación con la moral y la presencia del demonio. La desesperación o a la melancolía, así como sus remedios, estaban asociadas a lo religioso: la culpa y la tristeza como modo de expiación de los pecados. Por otro lado, el dolor derivado de los acontecimientos negativos se abordaba desde la fe en los designios divinos y en la importancia de la vida después de la muerte (Conti, 2007).

Durante la edad media los autores árabes recuperaron la tradición clásica e incorporaron causas de la melancolía relacionadas con el contenido de las preocupaciones, el miedo o la ira. Se profundizó en el papel del cerebro en la aparición de la melancolía (Avicena, siglo X) y se habló de elementos psicológicos, como la visión irreal de las cosas o el papel de la ociosidad en su desarrollo (Razes XI) (Conti, 2007).

En el año 1621, Robert Burton publicó una voluminosa obra que recogía la información aportada por los diferentes autores hasta esa fecha (Burton, 1621/2008). Para Burton, la melancolía comprendía un extenso grupo de dificultades que van desde la tristeza y los temores imaginados, a las alucinaciones y los delirios; provocado por causas endógenas (humorales) y exógenas (mala alimentación, soledad, ociosidad, causas divinas). Describía a los pacientes melancólicos como delgados, con mal aspecto físico, con el semblante triste y

“debido a sus grandes temores, penas y vejaciones, están embotados, pesados, perezosos, inquietos, incapaces de ponerse con ninguna ocupación”, se sienten culpables y les preocupa “el pasado, el presente y el futuro” (Burton, 1621/2008, p. 209). Añade que tienen graves problemas de sueño, aumento del pulso cardíaco y preocupaciones hipocondríacas, pero destaca que la variedad de síntomas es enorme entre personas y que ni siquiera la tristeza o el temor están presentes en todos los melancólicos, aunque si están presentes durante el tiempo suficiente, la melancolía llega seguro (Burton, 1621/2008).

No será hasta el siglo XVII cuando se rechace de manera definitiva la teoría de los humores y se atribuya la causa de la melancolía a los procesos químicos en el cerebro y en el corazón (indicando la importancia de las emociones). También es en este siglo se señala la importancia de los nervios y su papel en los pensamientos y delirios (Jackson, 1989).

El descubrimiento de la electricidad, permite a los investigadores teorizar sobre que una menor energía en la transmisión de los impulsos en el cuerpo provocaría un menor tono emocional (Willen Cullen, siglo XIX), desplazando definitivamente la atención al cerebro. Este autor realizó una clasificación de los “enfermedades nerviosas”, entre las que incluía la melancolía, en la que se englobaban sintomatologías muy diversas (Conti, 2007).

Una de las clasificaciones que más influyó en la identificación y separación de los procesos depresivos fue la desarrollada por Pinel a principios del siglo XIX, que distinguió entre manía, melancolía, idiocia y demencia. En esta distinción se reserva la categoría melancolía para las enfermedades que cursan con un “delirio parcial del pensamiento centrado en un objeto”, la preocupación constante por un elemento que se observa de modo desajustado. Su discípulo más influyente, Esquirol, dará un paso más al diferenciar las melancolías con sintomatología

centrada en la tristeza. Ambos autores destacan la influencia de las condiciones sociales y ambientales en el estado de ánimo, enfocando su atención hacia las emociones del paciente (Jackson, 1989). Desde que Esquirol escribiera su tesis denominada “Las pasiones como causas, síntomas y significados en el tratamiento de las enfermedades mentales” hasta la actualidad, una parte de los trastornos mentales se ha centrado en el estado emocional del paciente naciendo de manera más definida el estudio de los trastornos depresivos.

El primer autor que utilizó la palabra depresión fue Blackmore en 1715 (Conti, 2007), pero fue en el siglo XIX cuando se generalizó su uso (Berrios, 2008). La palabra melancolía, según señalaba Esquirol, ya no era apropiada para el uso científico, había de reservarse para un lenguaje más coloquial, para así evitar la confusión que desde Hipócrates hasta ese momento había entre las monomanías, los delirios y otros trastornos como la hipocondría. Además, el término melancolía se utilizaba por la población como sinónimo de tristeza, añoranza o carácter de ánimo negativo. De modo que Esquirol pasó a denominar a los periodos prolongados de tristeza con el nombre de lipemanía (Goldney, 2006) que seguía incluyendo muchos conceptos relacionados con los delirios y la depresión grave. Berrios (2008, p. 375) señala que: “Hasta el periodo de las Guerras Napoleónicas, la melancolía era un cajón de sastre de los problemas mentales cuyo denominador común era la presencia de pocas (en oposición a muchas) alucinaciones... La tristeza y el bajo estado de ánimo no eran considerados síntomas definitorios”, por lo que las depresiones no psicóticas no habían sido llamadas “melancolía en absoluto”.

El autor que supuso una mayor revolución en los sistemas clasificatorios de las enfermedades mentales fue Emile Kraepelin (1856-1926). A partir de una perspectiva descriptiva, diferenció entre la demencia precoz (posteriormente

denominada esquizofrenia) y la enfermedad maniaco-depresiva (caracterizada por cambios “excesos de afectividad”, recogiendo la idea ya avanzada durante el siglo XIX de que las personas que padecían estados depresivos alternaban su estado con periodos maniacos) (Millon, 2008).

Dentro de los *estados depresivos*, Kraepelin distinguió tres tipos por su gravedad: retardación simple, retardación con delirio y alucinaciones, así como los estados de estupor. La retardación simple correspondería con el actual episodio depresivo, que según este autor se caracterizaba por dificultades de concentración, falta de interés por las cosas, enlentecimiento psicomotor, dificultad para tomar decisiones, dificultad para recordar las cosas, gesto indolente, cansancio, pensamientos negativos, pesimismo sobre el futuro, ideación suicida, bajo rendimiento profesional o académico y conciencia de problema (Jackson, 1989), síntomas muy coincidentes con los actuales (APA, 2002). Además, este autor avanzó en el estudio del inicio y curso de la enfermedad, así como de los factores de herencia genética. Respecto de los periodos depresivos, señaló que tanto el inicio como la recuperación son graduales y que la duración del problema es de entre pocos meses y un año.

La circularidad de los cuadros clínicos que alternaban estados depresivos y maniacos fue observada por primera vez a mediados del siglo XIX, denominada en el caso de Kraepelin en sus últimas publicaciones como maniaco-depresivo, pero no será hasta el año 1957 cuando Karl Leonhard separó los trastornos afectivos en bipolares y monopolares, propuesta que ha sido ampliamente aceptada.

En contraposición con el sistema de clasificación desarrollado por Kraepelin y el énfasis que otorgaba a las explicaciones biologicistas, Adolf Meyer (1866-1950) señalaba que estas categorías eran demasiado inclusivas, puesto que muchos pacientes que se atendían en las consultas no respondían a esos patrones

cíclicos. Aunque creía que las categorías diagnósticas podían ser útiles, señalaba que era necesario conocer la historia de vida del paciente para comprender las causas de la depresión. Los estados de ánimo anormales eran en parte productos de procesos de adaptación anormales a la realidad, por lo que en esos sistemas se introducían en las mismas categorías aquellos procesos categorizados por estados depresivos, pero de orígenes muy diferentes (Jackson, 1989).

A lo largo del siglo XX se produjeron largas discusiones sobre la frontera entre los estados “normales” de tristeza y la depresión patológica. También fue conflictiva la distinción entre la depresión derivada de situaciones contextuales (depresiones reactivas o neuróticas) y los estados de depresión sin causa exterior aparente (depresión endógena o melancólica) (Vázquez y Sanz, 2008). Estas discusiones y las confusiones terminológicas que provocaron, permanecen en la actualidad, pero sin duda uno de los hechos más relevantes acaecidos durante el siglo XX en relación con la conceptualización de los trastornos depresivos fue la aparición de clasificaciones diagnósticas de los trastornos mentales de amplio uso y razonable validación que propusieron una definición consensuada de los trastornos depresivos que, con muy ligeras modificaciones, se mantiene hasta nuestros días.

Efectivamente, desde finales del siglo pasado y, en particular desde la publicación de la tercera edición del Manual Diagnóstico y Estadístico de los Trastornos Mentales (DSM-III) de la Asociación Americana de Psiquiatría (APA, 1980), el concepto de trastornos depresivos ha estado marcado por la hegemonía de las clasificaciones nosológicas, las cuales han tratado de consensuar, operativizar y dar un orden a las observaciones, conceptos y cuadros clínicos que se habían ido creando a lo largo de la historia. En particular, a partir de los años 70-80 del siglo pasado, las clasificaciones de trastornos mentales creadas desde una perspectiva atórica como la del DSM-III y sus sucesivas revisiones o la de la

Clasificación Internacional de Enfermedades (CIE) de la Organización Mundial de la Salud, han facilitado la comprensión de la salud mental en general y, de los trastornos depresivos en particular, al crear un lenguaje común para todas las orientaciones teóricas sobre la psicopatología.

### **1.1.2. Definición y clasificación de los trastornos depresivos**

Actualmente, los conceptos de trastorno depresivo más consensuados y validados, aunque no lejos de polémica y discusión, son los propuestos por la cuarta edición del Manual Diagnóstico y Estadístico de los Trastornos Mentales (DSM-IV) de la Asociación Americana de Psiquiatría (APA, 1994, 2002) y la clasificación de trastornos mentales y del comportamiento de la décima revisión de la Clasificación Internacional de Enfermedades (CIE-10) de la Organización Mundial de la Salud (OMS, 1993), los cuales distinguen a los trastornos depresivos de los trastornos del estado de ánimo que incluyen la manía o hipomanía (trastornos bipolares) y, a su vez, dentro de los trastornos depresivos distinguen al trastorno depresivo mayor y al trastorno distímico. Además, los conceptos de trastorno depresivo del DSM-IV y la CIE-10 guardan muchas semejanzas entre sí respecto a los síntomas y parámetros de duración, gravedad, curso e historia natural que definen el trastorno depresivo mayor y la distimia (Sanz, Izquierdo y García-Vera, en prensa).

En ambos tipos de trastornos depresivos, el núcleo de la definición de la depresión clínica es la perturbación prolongada del estado de ánimo negativo y la anhedonia. Los trastornos depresivos se caracterizan por la aparición de múltiples síntomas (APA, 2002; NICE, 2010; Sanz y Vázquez, 2008):

- **Síntomas emocionales:** la tristeza es la emoción central de los procesos depresivos, aunque en ocasiones los sujetos muestran síntomas de irritabilidad, desesperanza, angustia, ansiedad o sensación de vacío.

Otro síntoma central de los trastornos depresivos es la apatía y la anhedonia (incapacidad para experimentar placer, pérdida de interés o satisfacción en casi todas las actividades).

- **Síntomas somáticos:** durante los episodios depresivos aparecen con frecuencia alteraciones del sueño (insomnio, hipersomnia, despertares nocturnos, sensación de no haber descansado bien o somnolencia diurna) y del apetito (disminución o aumento de ingesta de comida o cambios en el peso corporal). También es habitual un aumento de la percepción de dolor, junto con una preocupación recurrente por el estado de salud, que con frecuencia es el motivo de consulta de los pacientes deprimidos. Existe a menudo una disminución en la libido del paciente y del interés por las relaciones sexuales.

- **Síntomas conductuales y motivacionales.** Un síntoma muy frecuente en pacientes deprimidos es la disminución de la cantidad de actividades que anteriormente realizaba. Pueden aparecer dificultades para levantarse de la cama por la mañana o para mantener las rutinas diarias más allá de la sensación de cansancio físico. Se puede mostrar abatido a nivel motor (retardo psicomotor) o por el contrario inquieto y con movimientos continuos. Además, aparecen signos externos de bajo estado de ánimo: llanto, mirada baja, bajo tono de voz o aspecto descuidado.

- **Cambios en el procesamiento cognitivo.** Las personas con sintomatología depresiva tienen pensamientos relacionados con la culpa, con el sentimiento de merecer ser castigado, baja autoestima, desesperanza o sentimientos de



inutilidad. Pueden pensar que la vida ya no merece la pena, manifestar pesimismo por el futuro, “visión en túnel” e ideación suicida (incluido intentos suicidas). Los pensamientos negativos se vuelven recurrentes durante la depresión y ocupan una parte importante de su actividad cognitiva. Aparecen también con frecuencia problemas de concentración o en la toma de decisiones, así como sesgos atencionales y atribucionales.

También se han descrito dificultades en la memoria, en la calidad de la atención, así como dificultades para el esfuerzo y el control ejecutivo (Vázquez y Sanz, 2008).

- **Aislamiento social e hipersensibilidad social.** Un 70% de los sujetos deprimidos dice haber disminuido su interés por la gente (Beck, 1976, en Vázquez y Sanz, 2008). El apoyo social percibido disminuye y los sentimientos de soledad son más frecuentes. Las preocupaciones recurrentes a menudo hacen que el paciente centre las conversaciones con los demás en sus problemas, sus quejas o en aspectos negativos de la realidad.

En casos graves, la sintomatología citada puede venir acompañada de síntomas psicóticos como delirios o alucinaciones. La depresión es un trastorno heterogéneo, con una gran variabilidad de síntomas entre distintos sujetos, con una gran comorbilidad con otros trastornos, que varían en un continuo de gravedad entre los trastornos clínicamente no significativos y los trastornos afectivos más graves.

La depresión es un problema egodistónico y tremendamente incapacitante. Con frecuencia el paciente incorpora la propia conciencia del problema a los pensamientos de preocupación y el abatimiento, aumentando los sentimientos de culpabilidad y baja autoestima.

Cuando esos síntomas que se han descrito responden a ciertos parámetros de duración, gravedad, curso o historia natural y, además, se han descartado algunas posibles causas, se habla de trastornos depresivos. En concreto, los parámetros y causas descartables que forman parte de las actuales definiciones de trastornos depresivos son aquellos que recogen los dos sistemas de clasificación y diagnóstico más utilizados actualmente en salud mental: el del DSM-IV, desarrollado por la Asociación Americana de Psiquiatría (APA, 2002), y la CIE-10, elaborada por la Organización Mundial de la Salud (OMS, 1993).

Puesto que el objetivo principal de esta tesis doctoral es valorar la eficacia diagnóstica<sup>1</sup> de las escalas de sintomatología depresiva del MCMI, las cuales han sido desarrolladas tomando como referencia distintas versiones del DSM, será el sistema de clasificación de este manual, en particular la de su última versión, el DSM-IV, el que se utilice como referencia en la presente tesis doctoral. Además, la mayoría de las publicaciones consultadas sobre el MCMI utilizan ese manual diagnóstico como referencia.

---

<sup>1</sup> En la literatura científica sobre la validez diagnóstica de los instrumentos diagnósticos, se ha utilizado de forma indistinta los términos “eficacia diagnóstica” (diagnostic efficacy), “rendimiento diagnóstico” (diagnostic performance) y “eficiencia diagnóstica” (diagnostic efficiency), aunque quizás este último sea el que con mayor frecuencia se utiliza. Sin embargo, dado el estado actual de la investigación sobre la validez y calidad de los instrumentos y pruebas diagnósticas, en la cual se distinguen diversas perspectivas, criterios y métodos para valorar dicha validez y calidad, parece necesario utilizar distintos términos para hacer alusión a esa diversidad teórica y metodológica y, por tanto, parece más apropiado decantarse por alguno de ellos para evitar confusiones. Sin pretender ser exhaustivos, parte de esa diversidad teórica y metodológica en el campo de los instrumentos y pruebas diagnósticas se ha reflejado, siguiendo la distinción ya consolidada en la investigación sobre intervenciones y tratamientos en salud (Cochrane, 1972; Haynes, 1999; Labrador, Echeburúa y Becoña, 2000; Saladrigas y Sacristán del Castillo, 2004), en la diferenciación de los términos “eficacia diagnóstica”, “efectividad diagnóstica” y “eficiencia diagnóstica” (Sox, Stern, Owens y Abrams, 1989; Stengel y Porzsolt, 2006). En este sentido, la eficacia diagnóstica haría alusión, por ejemplo desde una perspectiva médica, a “la correlación entre los hallazgos de la prueba y la verdad biológica (es decir, los hallazgos del patrón de oro diagnóstico)”, a la capacidad de un instrumento o prueba para “distinguir un paciente enfermo de un individuo sano” (Stengel y Porzsolt, 2006, p. 218), en definitiva, desde una perspectiva más amplia, al grado en que un instrumento o prueba mide lo que quiere medir, diagnóstica lo que quiere diagnosticar, discrimina entre un trastorno y la ausencia del mismo o la presencia de otro trastorno distinto, es decir, su validez diagnóstica. De este modo, los índices que pretenden comparar los resultados de un instrumento o prueba con un patrón oro diagnóstico (sensibilidad, especificidad, valor predictivo positivo, kappa, etc.) serían índices de validez diagnóstica o de eficacia diagnóstica. Por su parte, la efectividad diagnóstica haría referencia a la utilidad del instrumento o prueba diagnóstica en la práctica clínica, mientras que la eficiencia diagnóstica haría alusión a la relación costes-beneficios de la utilización de dicho instrumento o prueba diagnóstica. Finalmente, el término “rendimiento diagnóstico” tendría un significado más general y abarcaría tanto cuestiones de validez (eficacia) como de utilidad clínica (efectividad) o relación costes-beneficios (eficiencia). En la presente tesis doctoral, se utilizará pues, preferentemente, el término “eficacia diagnóstica” para referirse a la validez diagnóstica de las escalas del MCMI-II y de otros instrumentos psicopatológicos.

Entre las grandes categorías de trastornos mentales que propone el DSM-IV y, que dentro de su sistema de evaluación multiaxial, se evalúan en el Eje I (trastornos clínicos), se encuentran los trastornos depresivos del estado de ánimo (APA, 2002). El manual incluye dentro de esta categoría los cuadros sintomatológicos que tienen como “característica principal una alteración del humor”, que manifiestan una variación en el mantenimiento del estado afectivo caracterizado por la tristeza o la apatía, así como la euforia o la manía. Dentro de los trastornos de estado de ánimo, se encuentran dos categorías principales: los trastornos depresivos y los trastornos bipolares (véase la Tabla 1.1).

**Tabla 1.1.** Trastornos del estado de ánimo depresivo y trastornos bipolares en el DSM-IV (APA, 2002)

Trastornos depresivos	Trastornos bipolares
- Trastorno depresivo mayor episodio único	- Trastorno bipolar I, episodio maniaco único
- Trastorno depresivo mayor recidivante	- Trastorno bipolar (episodio más reciente hipomaniaco, maniaco, mixto, depresivo o no especificado)
- Distimia	
- Trastornos depresivos no especificados	- Trastorno bipolar II
	- Trastorno ciclotímico
	- Trastorno bipolar no especificado

La pieza fundamental para entender el diagnóstico de los trastornos del estado de ánimo son los episodios afectivos, distinguiéndose cuatro tipos: episodio depresivo mayor, episodio maníaco, episodio hipomaniaco y episodio mixto. En la Tabla 1.2. se pueden observar los síntomas recogidos para definir el episodio depresivo mayor y la distimia según el DSM-IV (para un mayor detalle véase el

Anexo 1), los cuales, como se ha comentado previamente, son muy similares a los recogidos por la CIE-10 (para un mayor detalle véase el Anexo 2).

**Tabla 1.2.** Síntomas que incluyen los criterios diagnósticos de los trastornos depresivos del DSM-IV y de la CIE-10 (adaptado de Sanz, Izquierdo y García-Vera, en prensa)

Síntomas	DSM-IV (1994)		CIE-10 (1993)	
	Depresión Mayor	Distimia	Depresión Mayor	Distimia
<b>Anímicos</b>				
1. Estado de ánimo o humor deprimido*	X	X	X	X
13. Llanto fácil				X
<b>Motivacionales</b>				
2. Disminución-pérdida del placer o interés	X		X	X
<b>Físicos</b>				
3. Fatigabilidad o pérdida de energía o vitalidad	X	X	X	X
4. Aumento/disminución del peso o apetito	X	X	X	
5. Insomnio o hipersomnia o alteraciones del sueño	X	X	X	X
<b>Motores</b>				
6. Agitación o enlentecimiento psicomotor	X		X	
9a. Conducta o intentos suicidas †	X		X	
15. Aislamiento social				X
16. Disminución de la locuacidad				X

Síntomas	DSM-IV (1994)		CIE-10 (1993)	
	Depresión Mayor	Distimia	Depresión Mayor	Distimia
<b>Cognitivos</b>				
7. Sentimientos excesivos o inapropiados de inutilidad o culpa o	X		X	
8. Disminución de la capacidad para pensar, concentrarse o tomar	X	X	X	X
9b. Ideas recurrentes de muerte o suicidio †	X		X	
10. Pérdida de confianza y autoestima, y sentimientos de	X	X	X	X
11. Sentimientos de desesperanza o desesperación		X		X
12. Sentimiento de incapacidad para enfrentarse con las responsabilidades				X
14. Pesimismo sobre el futuro o cavilaciones sobre el pasado				X

† El criterio sintomático 9 de actos/ideación suicidas se ha considerado como dos síntomas, uno motor (9a: actos suicidas) y otro cognitivo (9b: ideación suicida).

\* En el DSM-IV también se incluye la posibilidad de que en niños y adolescentes el estado de ánimo pueda ser irritable.

Además del episodio depresivo mayor, los criterios necesarios para el diagnóstico de trastorno depresivo mayor en el DSM-IV incluyen la ausencia de episodios maníacos, mixtos o hipomaniacos, así como la ausencia de un trastorno esquizoafectivo o del espectro psicótico que expliquen mejor los síntomas del episodio depresivo (véase el Anexo 3).

Como se puede observar en la Tabla 1.2, en el caso del trastorno depresivo mayor existen algunas diferencias entre el DSM-IV y la CIE-10. Para cumplir los criterios de episodio depresivo según la CIE-10, se deben cumplir al menos dos de los tres síntomas considerados como clave, en lugar de uno de los dos que recoge el DSM-IV. En el caso de la CIE-10, además de la tristeza y la anhedonia, se señala

también como síntoma clave la disminución de la vitalidad. Por otro lado aunque ambos manuales señalan que en general es necesario que los síntomas persistan al menos durante dos semanas, la CIE-10 añade que “períodos más cortos pueden ser aceptados si los síntomas son excepcionalmente graves o de comienzo brusco” (CIE-10, p. X). Existe en cualquier caso un gran solapamiento de síntomas (véase la Tabla 1.2).

En el caso del trastorno distímico no se aplica el concepto de episodio. La distimia se caracteriza por una duración de la sintomatología más prolongada (2 años), pero que incluye un número de síntomas menor para la asignación del diagnóstico (véase el Anexo 4). La intención de los sistemas de clasificación es detectar con esta categoría a aquellos trastornos que sin llegar a cumplir los criterios necesarios para el episodio depresivo mayor, fluctúan en el tiempo de manera más persistente (en ocasiones durante una parte muy significativa de la vida del sujeto). Es posible que se superponga durante el desarrollo de un trastorno distímico un trastorno depresivo mayor añadido. El diagnóstico diferencial entre ambos trastornos se detallará en el capítulo 2, junto con los trastornos depresivos no especificados.

En el caso de la distimia, las diferencias entre los manuales del APA y de la OMS son más notables que en el caso del trastorno depresivo mayor. Analizando la información de la Tabla 1.2, se puede encontrar que los criterios sintomáticos de distimia incluidos en el DSM-IV están todos ellos presentes en la CIE-10, pero, sin embargo, este último sistema diagnóstico añade numerosos síntomas más.

En cualquier caso, teniendo en cuenta que la “congruencia entre el DSM y la CIE es notable” Goldney (2006, p. 876), se ha señalado que ambos sistemas podrían ser unificados para asegurar un acuerdo universal en la conceptualización de los trastornos depresivos, para ser utilizado en investigación y en la práctica

clínica. Sin embargo, aunque no es frecuente la comparación del diagnóstico realizado con ambos manuales, algunos autores han encontrado diferencias significativas entre ellos (como se puede observar en Philips y Henderson, 1991).

Profundizando un poco más en la definición y clasificación de los trastornos depresivos del DSM-IV, es importante subrayar que los criterios diagnósticos del trastorno depresivo mayor han sido ligeramente modificados desde la publicación del DSM-III (APA, 1980), primera ocasión en la que su definición fue ampliamente estudiada de modo sistemático a partir de los Criterios Diagnósticos de Investigación (*Research Diagnostic Criteria*; RDC) propuestos Spitzer, Endicott y Robins (1978). En el DSM-IV el cambio más importante fue la ampliación del periodo temporal de los síntomas del episodio depresivo mayor, que pasó de una semana a dos. También se realizó una modificación en la agrupación de los síntomas: en el DSM-III, los elementos referidos al bajo estado de ánimo y la desesperanza estaban unidos, mientras que en el DSM-IV se incluyeron como síntomas independientes. Además, en el manual DSM-III no se incluía el criterio de deterioro en el funcionamiento psicosocial para el diagnóstico de los trastornos del estado de ánimo, que se añadió como un intento para evitar los falsos positivos en el diagnóstico.

Durante los últimos años, la Asociación Americana de Psiquiatría (APA) ha estado desarrollando una nueva edición del DSM, el DSM-5. Habrá que esperar a su publicación a lo largo del año 2013 para saber, en relación con los trastornos depresivos, qué cambios, incluye el DSM-5 en comparación con el DSM-IV, así como para poder analizar los datos y argumentos que justifican esos cambios. Por ahora, parece interesante mencionar los tres temas principales que han sido objeto de discusión por parte de la comisión y grupos de trabajo encargados de los trastornos del estado de ánimo en el futuro DSM-5 (APA, 2011):

Se ha discutido si los trastornos del estado de ánimo se deberían separar definitivamente en dos categorías los trastornos depresivos y los trastornos bipolares, sin que ambos quedaran encuadrados dentro de una categoría común. La justificación principal para este cambio es que se han encontrado importantes diferencias en la etiología y en el curso de los trastornos que engloban que pesan más que las similitudes sintomatológicas. A pesar de esta modificación, se insiste en que no existen diferencias significativas entre los episodios depresivos de los trastornos bipolares y la depresión unipolar.

En relación con el trastorno depresivo mayor, se ha debatido sobre la posibilidad de eliminar la distinción entre trastorno depresivo mayor y trastorno distímico, tal y como está reflejada en el DSM-IV, puesto que algunos autores defienden que no hay datos empíricos de que los trastornos depresivos mayores de larga duración (al menos dos años) se distingan del trastorno distímico, como señalaba el DSM-IV, debido a que el sujeto ha cumplido los criterios de episodio depresivo mayor durante el primer año del inicio del trastorno.

Respecto a la definición de episodio depresivo mayor, se ha discutido la posible eliminación del criterio que afirma que, para poder diagnosticar dicho episodio, los síntomas no se deben explicar mejor por la presencia de un duelo. Este cambio se justificaría debido a la no existencia de datos empíricos que indiquen que los estados depresivos derivados de otros eventos estresantes sean distintos a los provocados por el duelo (Zisook y Kendler, 2007).

En cualquier caso, y a la espera de su publicación definitiva, es importante señalar que los cambios que se han debatido en relación con el DSM-5 no suponen ninguna modificación en la definición actual de los trastornos depresivos del DSM-IV respecto a los síntomas que definen dichos trastornos o a su duración y gravedad mínima.



### Curso de los trastornos depresivos

Los síntomas que conforman el episodio depresivo mayor suelen aparecer de manera progresiva. Según el DSM-IV “puede haber un período prodrómico con síntomas ansiosos y síntomas depresivos que puede durar semanas o meses” (APA, 1996, p. 331). Según la guía NICE (2010, p. 19) “otros sujetos pueden desarrollar un trastorno depresivo mayor grave considerablemente rápido, con frecuencia tras la aparición de un estresor vital significativo”. Se considera que “lo habitual es que un episodio no tratado dure 6 meses o más, independientemente de la edad de inicio” (APA, 2002, p. 331). “Aunque la depresión ha sido caracterizada como un trastorno limitado en el tiempo, con una duración de entre 4 y 6 meses, ahora tenemos claro que la remisión incompleta y la recaída son comunes” (NICE, 2010, p. 19). Se calcula que entre un 20% y 30% de casos se mantienen algunos síntomas depresivos “insuficientes para cumplir totalmente los criterios para un episodio depresivo mayor, pero que persisten durante meses o incluso años y se asocian a incapacidad o malestar... En algunos sujetos (5-10%) se siguen cumpliendo totalmente los criterios para el episodio depresivo durante 2 años o más” (APA, 2002, p. 331).

El trastorno depresivo mayor es un trastorno recurrente. El 72.3% de las personas con depresión mayor habían tenido más de un episodio previamente según el estudio NCS realizado en EE. UU. (Kessler et al., 1994). Sin embargo, otros estudios señalan tasas cercanas al 40% (Bijl, Ravelli y Zessen, 1998; Kessler, Berglund, Demler, y Jin 2003). Cuando los síntomas están asociados a trastornos de ansiedad parece haber una mayor probabilidad de estabilidad en el trastorno depresivo. En población clínica se ha encontrado una tasa de recurrencia del trastorno del 60% a los 5 años, 67% a los 10 años y 85% después de 15 años: en población general, en cambio, es más baja (35% después de 15 años) (Hardeveld,

Spijker, De Graaf, Nolen y Beekman, 2010). Estos autores encontraron que las variables clínicas que mejor predecían la recaída eran el número de episodios anteriores y la sintomatología residual de los sujetos (remisión parcial). Se ha señalado que después de cada episodio depresivo, “cantidades cada vez más pequeñas de estresores pueden disparar reacciones de una intensidad comparable” (Kendler et al., 2000, citado en Sanz y Vázquez, 2008, p. 24).

El trastorno distímico tiene “características asociadas del trastorno distímico son parecidas a las del episodio depresivo mayor” (APA, 2002, p., 354), pero con frecuencia tiene una edad de inicio más temprana y un curso crónico. Además, es habitual que las personas con un trastorno distímico tengan un episodio depresivo mayor y una mayor probabilidad de no tener una recuperación total tras el episodio y una mayor probabilidad de recaída (APA, 2002).g

### **Etiología de los trastornos depresivos**

Existen muchas preguntas que generan conflictos sobre las causas de la depresión (Beck y Alford, 2009). Algunos investigadores consideran la depresión como un trastorno psicológico primario; mientras que otros mantienen de manera igualmente firme que está causado por factores orgánicos. Existe una amplia gama de factores biológicos, psicológicos y sociales, que tienen un impacto significativo en el curso de la depresión y la respuesta al tratamiento. Aunque no se conocen exactamente las causas de este trastorno en cada persona, sí sabemos con certeza que la mayoría de los trastornos depresivos son síndromes multicausales. Desde una perspectiva biopsicosocial, se asume que los trastornos depresivos son producto de la interacción entre factores genéticos, ambientales, comportamentales, emocionales y cognitivos, que actúan de manera diferencial en cada paciente deprimido.

Desde el punto de vista **genético** se argumenta que según los estudios realizados con gemelos, entre el 40% y el 50% de la susceptibilidad a padecer un trastorno depresivo mayor es debido a la influencia de la genética según Levinson (2009) y entre el 30% y 40% según Ebmeier et al. (2006).

Un metaanálisis encontró una tasa de heredabilidad del trastorno depresivo mayor del 37%, independientemente de la gravedad del trastorno, poniendo en duda la creencia de que las formas más graves de depresión tienen un componente más biológico (Sullivan, Neal y Kendler, 2000, en Vázquez y Sanz, 2008).

Desde el punto de vista bioquímico, la depresión está provocada por factores orgánicos del sistema nervioso central (Ebmeier et al., 2006). Desde mediados de los años 50 del siglo pasado, la aparición de fármacos antidepresivos eficaces permitió a los investigadores descubrir cómo la modificación de los neurotransmisores en las hendiduras sinápticas provocaba aumentos en el estado de ánimo de los pacientes. En especial, la modificación de la transmisión neuronal de las vías serotoninérgicas y noradrenérgicas tiene un efecto antidepresivo en un alto porcentaje de pacientes. El hecho de que esos cambios se lleven a cabo desde la primera toma de la medicación y en cambio el estado de ánimo mejore gradualmente semanas después del inicio de la toma, ha generado una gran cantidad de descubrimientos sobre los mecanismos neurológicos que operan en el efecto de estos psicofármacos y por tanto, en el conocimiento sobre el funcionamiento de la depresión. La explicación biológica sobre el funcionamiento de la depresión ha estudiado componentes sinápticos, celulares, genéticos y anatómico/funciones, lo que ha permitido entender de un modo más completo las “los obstáculos reales a los que combaten los pacientes deprimidos” (Ebmeier, et al., 2006, p. 158).

En general, una vida con eventos estresantes no lleva necesariamente a tener episodios depresivos, pero las personas que experimentan sucesos estresantes antes de la aparición de un episodio depresivo sí muestran un peor funcionamiento y una mayor gravedad de los síntomas y, además, algunos sucesos estresantes, tales como algunos que ocurren durante la infancia (p. ej., muerte o divorcio de los padres), están asociados con un mayor riesgo de padecer depresión (Colman y Ataullahjan, 2010). Es decir, los eventos estresantes de la infancia parecen tener un papel importante, pero su influencia no es decisiva. En cambio los eventos estresantes más cercanos en el tiempo, como la pérdida de un empleo o un divorcio (en general la pérdida de una persona significativa), aumentan la probabilidad de que una persona de tener depresión y de tener sintomatología más intensa. Acontecimientos como la muerte de una persona cercana, la pérdida de trabajo, la situación de desempleo o los eventos traumáticos han sido asociados repetidamente con episodios depresivos (Colman y Ataullahjan, 2010).

Las teorías conductistas señalan que la pérdida de reforzadores y, sobretudo, la pérdida de eficacia en obtenerlos, produce un bajo estado de ánimo, que disminuye a su vez la capacidad de afrontamiento del sujeto (Dimidjian, Barrera, Martell, Muñoz, y Lewinsohn, 2011). El origen de este proceso podía tener relación tanto con déficits internos del sujeto, como con cambios en las circunstancias ambientales tales como los producidos por los acontecimientos estresantes a los que se aludía antes.

La terapia cognitiva explica el origen de la depresión con el modo peculiar que tiene el individuo de verse a sí mismo, su entorno y su futuro y en su forma idiosincrásica de procesar la información (Beck, Rush, Shaw y Emery, 2005).

Vázquez, Hervás, Rahona y Gómez (2009), en su revisión sobre los modelos cognitivos de la depresión, han encontrado un apoyo empírico claro para

el establecimiento de los sesgos cognitivos como variable de mantenimiento y pronóstico del curso de los trastornos depresivos. Además, aspectos como la inestabilidad de la autoestima, las situaciones vitales estresantes, los episodios depresivos previos, los estilos atribucionales, los sesgos de memoria, los sesgos atencionales y el estilo rumiativo de afrontamiento forman parte de las variables de vulnerabilidad para la depresión.

Es difícil discriminar entre las causas, los síntomas y las consecuencias de los trastornos depresivos. Por ejemplo, para algunos autores las dificultades en las relaciones sociales de los sujetos con depresión no son más que una consecuencia del proceso depresivo endógeno, mientras que para otros es precisamente su causa o su factor mantenedor. En algunos modelos psicológicos los pensamientos rumiativos se pueden considerar como mantenedores del estado depresivo, mientras que para otros son tan solo una manifestación más de una disfunción en el funcionamiento de los neurotransmisores.

### **1.2. Epidemiología e importancia sociosanitaria de los trastornos depresivos**

Los estudios epidemiológicos de las últimas décadas han proporcionado información fundamental sobre los trastornos depresivos. Ahora se tiene evidencia de que tales trastornos se encuentran entre los problemas más importantes para la salud pública a nivel mundial porque tienen una alta prevalencia y producen un gran malestar y un deterioro muy importante del desempeño social, laboral, etc. (Kessler, 2007; Knapp, 2006; Üstün, Ayuso-Mateos, Chatterji, Mathers y Murray, 2004).

El estudio epidemiológico de los trastornos mentales se ha desarrollado de manera sistemática a partir de la creación de los criterios diagnósticos y de las entrevistas estructuradas a mediados de la década de los años 70 del siglo pasado.

Existen antecedentes para estos estudios que se pueden dividir en las siguientes etapas (Goodwin, et al., 2006):

- Desde mediados del siglo XIX, los investigadores recogieron información proporcionada por los profesionales que atendían a los pacientes (médicos, enfermeros o clérigos). La información no se integraba en un sistema diagnóstico, sin embargo las descripciones sintomáticas eran progresivamente más detalladas.
- A principios del siglo XX se inició la recogida de información directamente de los pacientes, mediante entrevistas y cuestionarios aplicados por especialistas en salud mental. En ese momento no existían criterios uniformes para el diagnóstico y no se habían desarrollado las entrevistas estructuradas, por lo que la fiabilidad interjueces era baja. Los criterios utilizados incluían los trastornos depresivos leves como neurosis, encuadrados junto con los trastornos de ansiedad, mientras que los trastornos bipolares y los trastornos psicóticos se recogían dentro de las depresiones psicóticas.
- La elaboración de criterios diagnósticos más explícitos a mediados de los años 70 del siglo XX permitió el desarrollo de instrumentos de medida más válidos y eficaces. Esta corriente se concretó en la elaboración de los manuales DSM-III (APA, 1980), DSM-IV (APA 1994), CIE-9 (OMS, 1977) y CIE-10 (1992). En paralelo a estas guías, se crearon entrevistas estructuradas para detectar la presencia de los distintos trastornos en la población con una mayor fiabilidad. Las principales conclusiones a las que se han llegado a partir de estos avances instrumentales son que los trastornos mentales son comunes entre la población y que la mayor parte de las personas que los padecen no acceden a los tratamientos adecuados.

Antes de comentar los resultados de los principales estudios epidemiológicos en lo que concierne a los trastornos depresivos, es importante aclarar algunos conceptos necesarios para su interpretación:

- **Incidencia:** se define como el número de casos nuevos de una enfermedad o trastorno en una población y un periodo de tiempo determinados. Por tanto nos transmite información acerca del riesgo de padecer un episodio nuevo del trastorno.
- **Prevalencia:** es el número de casos de una enfermedad o trastorno en una población y en un momento dado. Define por tanto al porcentaje de sujetos que padecen una enfermedad o trastorno con respecto a la población. En particular, en los estudios de los trastornos mentales se utilizaran habitualmente datos de prevalencia a los 12 meses y de prevalencia vital (desde el nacimiento de la persona hasta el momento de la realización de la evaluación), con independencia de cuando se haya iniciado el trastorno. Algunos autores señalan que el hecho de preguntar acerca de trastornos depresivos en el pasado puede suponer un serio reto (Andrews et al., 1999; Pattern, 2003; Wells y Horwood, 2004, en Slade y Sunderland, 2010). Por ejemplo, se considera “particularmente problemático evaluar las tasas de prevalencia vital. Las estimaciones obtenidas mediante una sola entrevista son problemáticas teniendo en cuenta que el recuerdo tiende a ser incompleto y selectivo (Paykel, 2000, en Slade y Sunderland, 2010).
- La **prevalencia vital proyectada** es una estimación del porcentaje de personas que tendrán el trastorno a lo largo de toda su vida (Kessler et al., 2007). Existen dudas acerca de la utilidad de esta estimación. Por ejemplo, en un estudio sobre población canadiense se observó que utilizando una metodología longitudinal, en 6 años de incidencia acumulada, la prevalencia

encontrada era mayor que con el cálculo de la prevalencia vital estimada (Wang et al., 2010).

### 1.2.1. Prevalencia de los trastornos depresivos

A continuación se revisarán los resultados de los estudios epidemiológicos más importantes realizados hasta la fecha respecto a la prevalencia de los trastornos mentales y, en particular, la prevalencia de los trastornos depresivos.

#### Los estudios epidemiológicos en EE. UU. con entrevistas estructuradas

Para llevar a cabo el primer estudio nacional en EE. UU., denominado ECA (*Epidemiologic Catchment Area*), se aplicó la entrevista DIS (Robins, Helzer, Croughan y Ratcliff, 1981) para realizar diagnósticos DSM-III. La prevalencia vital del conjunto de los trastornos del estado de ánimo fue de 8.3% y la prevalencia a los 12 meses de 6.3%. En el caso del trastorno depresivo mayor, la prevalencia vital fue de 5.9% y de 3.5% durante los últimos 12 meses (Bourdon, Rae, Locke, Narrow y Regier, 1992).

Poco después se publicaría un estudio realizado mediante la entrevista estructurada CIDI (OMS, 1990) denominado *National Comorbidity Survey* (NCS), que indicaba unas tasas de prevalencia del trastorno depresivo mayor más altas: en concreto del 10.3% a los 12 meses, siendo la prevalencia vital de 17.1%. Para el conjunto de los trastornos del estado de ánimo la prevalencia a los 12 meses fue de 11.3% y para la prevalencia vital de 19.3%. En este estudio se publicó una prevalencia de la distimia del 2.5% a los 12 meses (Kessler et al., 1994). La interpretación de estas primeras encuestas nacionales fue que los trastornos mentales eran más frecuentes de lo esperado y que quizá los umbrales diagnósticos estaban demasiado bajos.



En la replicación de este último estudio denominado *National Comorbidity Survey Replication* (NCS-R), la prevalencia vital del trastorno depresivo mayor fue de 16.2% y la prevalencia a los 12 meses de 6.6%. Se utilizó de nuevo la entrevista CIDI, en este caso modificada para acomodarla a los criterios DSM-IV (Kessler et al., 2003), que incluyeron el criterio de significación clínica para que los evaluadores no sobrediagnosticaran a los sujetos. Los datos más actualizados están disponibles en Harvard School of Medicine (2007).

Estos datos son compatibles con un estudio posterior realizado sobre una amplia muestra de población estadounidense, en la que se encontró una prevalencia del trastorno depresivo mayor a los 12 meses de 5.28% y una prevalencia vital de 13.23% (Hasin, Goodwin, Stinson y Grant, 2005).

### **Iniciativa *World Mental Health* (WMH) de la Organización Mundial de la Salud (OMS)**

Se trata del mayor estudio epidemiológico de los trastornos mentales realizado hasta el momento, desarrollado por la OMS en un amplio número de países. A través de la entrevista CIDI Versión 3.0 (Kessler y Üstün, 2004), se recogieron datos de 17 países y se realizaron las entrevistas a 85 052 personas representativas de la población estudiada (en dos de los países se circunscribía a las poblaciones urbanas). Este proyecto utilizó la misma metodología en una amplia variedad de regiones, por lo que los resultados son fácilmente comparables. Se evaluaron las categorías diagnósticas de trastornos del estado de ánimo, ansiedad, control de impulsos y consumo de sustancias, para generar diagnósticos DSM y CIE, dejándose fuera del proyecto los trastornos psicóticos y los trastornos adaptativos.

Los datos mostraron que los países con altos ingresos tenían prevalencias ligeramente más altas de media (prevalencia vital 14.6%) que los países de bajos ingresos (prevalencia vital 11.1%) (Bromet et al., 2011). Sin embargo, en la prevalencia a los 12 meses las diferencias no fueron significativas (5.5% y 5.9% respectivamente) (Kessler et al., 2010). En la Tabla 1.3 se recogen resultados más detallados de este estudio respecto a la prevalencia vital y la prevalencia estimada de los diferentes grupos de trastornos evaluados, incluyendo los trastornos del estado de ánimo.

**Tabla 1.3.** Prevalencia vital y estimada a la edad de 75 años de los trastornos DSM-IV en la iniciativa WMH de la OMS (Kessler et al., 2007)

País	Ansiedad		Estado de ánimo		Control de impulsos		Consumo de drogas		Cualquier trastorno	
	Prev. vital	Prev.	Prev.	Prev.	Prev. vital	Prev.	Prev. vital	Prev.	Prev. vital	Prev.
		estimada	vital	estimada		estimada		estimada		estimada
Bélgica	13.1	15.7	14.1	22.8	5.2	5.2	8.3	10.5	29.1	37.1
Colombia	25.3	30.9	14.6	27.2	9.6	10.3	9.6	12.8	39.1	55.2
Francia	22.3	26.0	21.0	30.5	7.6	7.6	7.1	8.8	37.9	47.2
Alemania	14.6	16.9	9.9	16.2	3.1	3.1	6.5	8.7	25.2	33.0
Israel	5.2	10.1	10.7	21.2			5.3	6.3	17.6	29.7
Italia	11.0	13.7	9.9	17.3	1.7		1.3	1.6	18.1	26
Japón	6.9	9.2	7.6	14.1	2.8		4.8	6.2	18.0	24.4
Líbano	16.7	20.2	12.6	20.1	4.4	4.6	2.2		25.8	32.9
México	14.3	17.8	9.2	20.4	5.7	5.7	7.8	11.9	26.1	36.4
Holanda	15.9	21.4	17.9	28.9	4.7	4.8	8.9	11.4	31.7	42.9
N. Zelanda	24.6	30.3	20.4	29.8			12.4	14.6	39.3	48.6
Nigeria	6.5	7.1	3.3	8.9	0.3		3.7	6.4	12.0	19.5
China	4.8	6.0	3.6	7.3	4.3	4.9	4.9	6.1	13.2	18
Sudáfrica	15.8	30.1	9.8	20.0			13.3	17.5	30.3	47.5
España	9.9	13.3	10.6	20.8	2.3	2.3	3.6	4.6	19.4	29.0
Ukrania	10.9	17.3	15.8	25.9	8.7	9.7	15.0	18.8	36.1	48.9
EE. UU.	31.0	36.0	21.4	31.4	25.0	25.6	14.6	17.4	47.4	55.3

Como se puede observar en la Tabla 1.3, la prevalencia vital del conjunto de los trastornos variaba desde 47.4% en Estados Unidos hasta el 12% en Nigeria, con un rango intercuartíl (RIC; percentil 25-75) de 18.1% y 36.1% (Kessler et. al., 2007). Los autores destacan que dependiendo de las regiones, el porcentaje de sujetos que cumplían los criterios necesarios para algún trastorno variaba significativamente. Si en algunos países como EE. UU. casi la mitad de la población cumplía los criterios diagnósticos de un trastorno mental, este porcentaje se reducía a una sexta parte en países como España (19,4%). Mientras que el conjunto de los trastornos de ansiedad fueron los más prevalentes en 10 países (4.8-31.0%), los trastornos del estado de ánimo fueron los más prevalentes en 6 países (3.3-21.4%) (Kessler et al., 2007).

En el caso de la población española, más de un cuarto de la población (29%) tenía un trastorno mental, mientras que en países como EE. UU., aproximadamente la mitad de la población (Kessler et al., 2007). Como se puede observar en la Tabla 1.3, los trastornos depresivos tenían una prevalencia vital de 10.6% en España (la categoría con mayor prevalencia), datos similares a los encontrados en otros países europeos como Alemania e Italia, ambos con un 9.9% de personas, pero menor que en otros países como Francia (21%) o Estados Unidos de América (21.4%).

Los autores del estudio comentaban que la prevalencia podía estar subestimada debido a que las personas con trastornos mentales suelen rechazar con más frecuencia su participación en este tipo de estudios. La aceptación de los problemas mentales es distinta en cada país, lo que puede explicar las grandes diferencias de prevalencia entre los distintos países. Además, es posible que existan sesgos debido a que la dificultad para reconocer la sintomatología es mayor en los países menos desarrollados. Por otro lado, consideraban que podría haberse

sobreestimado la prevalencia debido a que se hubiesen incluido perfiles subclínicos como si tuvieran de hecho el trastorno (Kessler et al., 2007).

### Estudios epidemiológicos europeos de los trastornos depresivos

A nivel europeo, dentro de la iniciativa *World Mental Health* de la OMS, se han realizado estudios epidemiológicos en seis países (Bélgica, Francia, Alemania, Italia, Holanda y España). Este proyecto, denominado ESEMeD (Alonso et al., 2004a), es la iniciativa Europea más completa para el cálculo de la prevalencia de los trastornos mentales realizada hasta la fecha.

Según este estudio (véase la Tabla 1.4), uno de cada cuatro participantes de la muestra cumple los criterios diagnósticos de un trastorno mental a lo largo de su vida y una de cada 10 en los últimos 12 meses. El 14% ha padecido un trastorno del estado de ánimo durante su vida, mientras que durante los últimos 12 meses la prevalencia es de 4.2%. Específicamente, la depresión mayor tiene en este estudio una prevalencia vital de 12.8%, mientras que su prevalencia a los 12 meses es de 3.9%. En el caso de la distimia, la prevalencia vital fue de 4.1% y la prevalencia a los 12 meses 1.1% (Demyttenaere et al., 2004).

**Tabla 1.4.** Prevalencia vital y prevalencia a los 12 meses de los trastornos mentales y los trastornos del estado de ánimo en seis países Europeos (Demyttenaere et al., 2004)

ESEMeD	Total	Total
Cualquier trastorno mental	25.0	9.6
Trastornos del estado de ánimo	14.0	4.2
Trastorno depresivo mayor	12.8	3.9
Trastorno distímico	4.1	1.1

### **Otras publicaciones sobre la epidemiología de los trastornos depresivos en Europa**

El primer estudio trasnacional europeo específico sobre la depresión se denominó DEPRESS (*Depression Research in European Society*). En este estudio se calculó la prevalencia para los 6 países mencionados anteriormente (Bélgica, Francia, Alemania, Italia, Holanda y España), sobre una muestra de 78 453 personas. La prevalencia de la depresión durante los últimos 6 meses fue del 17%, siendo la prevalencia de la depresión mayor un 6.8%. La diferencia entre estas dos cifras se debe a que se detectó un 8.3% de los sujetos que presentaban sintomatología depresiva, pero sin cumplir los criterios de deterioro funcional (Lépine, Gastar, Mendlewicz y Tylee, 1997).

Por su parte, el proyecto ODIN (*Outcome of Depression International Network*) nació de la Unión Europea en el año 1996 con el objetivo de estimar la prevalencia y los factores de riesgo de los trastornos depresivos y medir el impacto de dos tipos de intervenciones en este tipo de trastornos (Ayuso-Mateos et al., 2001). Para lograrlo se evaluó mediante la entrevista SCAN (Wing et al., 1990) a aquellos sujetos que habían obtenido una puntuación mayor de 12 en el BDI. Se estudiaron las prevalencias en 5 países: Reino Unido, Finlandia, Noruega, Irlanda y España. En el global de los 5 países se encontró un 8.56% de personas con trastornos depresivos durante el último mes. En el caso de España los datos fueron obtenidos a partir de una muestra de personas de Cantabria y la prevalencia estimada fue de 2.6% (Ayuso-Mateos et al., 2001).

Wittchen y Jacobi (2005) realizaron una selección sistematizada de estudios disponibles en Europa que cumplieran con unas condiciones estándar de calidad metodológica (muestreo aleatorio, datos de prevalencia puntual o a los 12 meses, algoritmos diagnósticos apropiados,...). Los resultados de la revisión de Wittchen y

Jacobi (2005) mostraban que en los 28 países europeos estudiados, el trastorno más frecuente en los sujetos desde los 18 a los 65 años fue el trastorno depresivo mayor (media = 6.9%), por encima de las fobias específicas (media = 6.6%) y los trastornos somatomorfos (media = 6.3%). En el caso del trastorno depresivo mayor, los autores destacaban la consistencia en los datos encontrados, con un rango de prevalencias entre 3.1% y 10.1% y un alto porcentaje de estudios situado muy próximos a la media. Aquellos estudios que utilizaron entrevistas clínicas realizadas por especialistas con experiencia solían obtener tasas de prevalencia menores que aquellos que utilizaban personal no clínico para realizar las entrevistas como la CIDI, dato que no ha sido confirmado en otros estudios (Haro et al., 2006b).

Wittchen y Jacobi (2005) señalan que existen muchas limitaciones para el cálculo de la prevalencia de los trastornos mentales en Europa. La principal dificultad es que existen pocos países que de manera consistente realicen encuestas nacionales que tengan como objetivo recoger datos epidemiológicos sobre salud mental a nivel nacional. Solamente Inglaterra, Alemania y Holanda llevan a cabo encuestas de ese tipo y merece la pena describir brevemente sus resultados principales.

En **Inglaterra**, la Encuesta Nacional de Morbilidad Psiquiátrica (*National Psychiatric Morbidity Survey*) se ha realizado en tres ocasiones, en 1993, 2000 y 2007 (Jonas et al., 2011). En estos estudios epidemiológicos se ha aplicado la entrevista CIS-R (*Revised Clinical Interview Schedule*; Lewis, Pelosi, Araya y Dunn, 1992) para el diagnóstico de los trastornos mentales según criterios CIE-10.

Siguiendo la metodología de los estudios anteriores con el objetivo de poder comparar los resultados entre las distintas encuestas, en la última versión del estudio realizado hasta la fecha los autores valoraron en la muestra la prevalencia durante la última semana de los distintos trastornos. El episodio depresivo mayor

tuvo una prevalencia del 2.3%, mientras que el trastorno más prevalente fue el trastorno mixto ansiedad-depresión, con 9% (Jonas et al., 2011). La prevalencia de los episodios depresivos fue del 2.1% en el estudio anterior (Jenkins et al., 2003) y del 2% en el primer estudio (Jenkins et al., 1997). Por tanto, los autores señalaban que se habría detectado un ligero incremento en la prevalencia de los trastornos mentales desde la primera a la última encuesta (Jonas et al., 2011).

En el caso de **Holanda**, se llevó a cabo el proyecto NEMESIS (Encuesta Nacional de Salud Mental e Incidencia). Existen dos estudios publicados hasta la fecha (Bijl, et al, 1998; de Graaf, ten Have, van Gool y van Dorsselaar, 2012), y en ambas encuestas se entrevistó a personas con edades entre los 18 y 64 años, mediante la entrevista CIDI 1.1 en el primer caso y CIDI 3.0 en el segundo (Kessler y Üstün, 2004).

Ambos estudios obtuvieron resultados similares con respecto a los trastornos del estado de ánimo. Los datos del estudio NEMESIS 2 señalan que el 44.7% de la población por debajo de 65 había padecido un trastorno mental a lo largo de su vida y el 17.7% lo tuvo durante el último año. En el conjunto de los trastornos del estado de ánimo, se encontró una prevalencia vital del 20.2%, siendo la prevalencia durante el último año de 6.1%. Específicamente en la depresión mayor, la prevalencia vital fue del 18.7% y a los 12 meses fue del 5.2%. El caso de la distimia la prevalencia vital fue del 1.3% y la prevalencia al año del 0.9%, encontrándose aquí una gran diferencia con respecto al estudio anterior en el que la distimia tenía una prevalencia vital de 6.3% y una prevalencia al año de 2.3% (de Graaf et al., 2011).

La Encuesta del Servicio de Salud Nacional de **Alemania** (GHS), realizada en una muestra de 4 181 sujetos de entre 18 y 65 años, señala una prevalencia a los 12 meses de los trastornos mentales del 31% y una prevalencia vital del 43%.



Los trastornos depresivos (trastorno depresivo mayor junto con trastorno distímico) tienen una prevalencia de 10.7% a los 12 meses, con una prevalencia vital del 17.1% (Jacobi, et al., 2004).

En el caso de Alemania, se pueden comparar los datos de esta encuesta con los proporcionados por el proyecto ESEMeD. Se encuentran diferencias muy significativas en el caso de la prevalencia vital de los trastornos del estado de ánimo. Mientras que en el estudio de la OMS los datos señalan un 9.9%, en el caso de la encuesta GHS señalan un 18.6%. En España en cambio la encuesta ESEMeD es la única referencia nacional de la que se dispone.

### **Prevalencia de los trastornos depresivos en España**

El estudio de la iniciativa *World Mental Health* (estudio ESEMeD) en España se llevó a cabo durante los años 2001 y 2002, con una muestra de 5 473 personas seleccionadas aleatoriamente en el territorio nacional (Haro et al., 2006a).

El porcentaje de personas que presentaron un trastorno mental en algún momento de su vida fue del 19.46%, mientras que durante los últimos 12 meses fue del 8.4% (véase la Tabla 1.5.). En el caso de los trastornos depresivos, existe una confusión en cuanto a la prevalencia vital porque según el artículo publicado por Kessler et al. (2007) el 10.6% de la población ha cumplido los criterios de algún trastorno del estado de ánimo, mientras que Haro et al. (2006a) señalan un porcentaje del 11.47%. Cabe la posibilidad de que se deba a un error tipográfico, ya que en el caso de Kessler et al. (2007) la prevalencia anotada puede ser en realidad la correspondiente al trastorno depresivo mayor, porque coincide con el porcentaje reflejado para este trastorno en Haro et al. (2006a). En cuanto a la prevalencia durante los últimos 12 meses, el 4.3% de la población mostraba al menos un trastorno del estado de ánimo y un 3.9% un trastorno depresivo mayor

(Haro et al, 2006a). El riesgo de prevalencia vital estimada, antes de los 75 años, de los trastornos depresivos es de 20.8%, siendo significativamente superior al del conjunto de los trastornos de ansiedad (13.3%) (Kessler et al., 2007).

Podemos observar que la depresión mayor es el trastorno más prevalente, tanto en la prevalencia a los 12 meses como en la prevalencia vital. Los datos indican una menor prevalencia en el caso de la distimia (1.49% durante el último año). Los autores destacan el carácter episódico de los trastornos depresivos, frente al carácter más crónico de los trastornos de ansiedad (Haro et al., 2006a). Según los resultados de este estudio, las tasas de prevalencia de depresión mayor en España son menores que en otros países europeos y similares a los de otros países del sur de Europa como Italia (Gabilondo et al., 2010).

**Tabla 1.5.** Prevalencia de los trastornos mentales y los trastornos depresivos en España según el estudio ESEMeD (Haro et al., 2006a)

ESEMeD España	Prevalencia Vital			Prevalencia 12 meses			Prev. vital estimada
	Total	Hombres	Mujeres	Total	Hombre	Mujeres	Total
Cualquier trastorno	19.4	15.7	22.9	8.5	5.2	11.4	29.0
Trastornos depresivos	11.4 <sup>a</sup>	6.7	15.8	4.3	6.2	2.3	20.8
Trastorno depresivo mayor	10.6	6.4	14.5	4	2.2	5.4	No disponible
Trastorno distímico	3.6	1.8	5.3	1.5	0.5	2.4	No disponible

<sup>a</sup> Según Haro et al., 2006a (10.6 en Kessler et al., 2007a)

A parte del estudio realizado dentro de la encuesta ESEMeD, los estudios epidemiológicos llevados a cabo en España se han basado en muestras pequeñas y poco representativas del estado (Haro et al., 2006a; Gabilondo et al., 2010):

- En una muestra de población del País Vasco con individuos entre los 18 y los 64 años, se obtuvieron unos datos de prevalencia vital de 20.6% en varones y 36.7% en mujeres (Gaminde, Uria, Padro y Querejeta 1993) datos muy superiores a los del estudio ESEMeD.
- Utilizando criterios del manual DSM-III, se obtuvo una prevalencia puntual de depresión de 4.3% en varones y un 7.8% en mujeres (Vázquez-Barquero et al., 1987, en Haro et al., 2006a).
- Sobre una muestra de la población mayor de 15 años de la isla de Formentera, que encontró una prevalencia puntual de trastornos mentales de un 21.4% y para los trastornos del estado de ánimo un 3.4%. Se utilizó en el estudio la entrevista estructurada SCAN y el cuestionario GHQ-28 (Roca et al., 1999; Roca, Gili, Ferrer y Bernardo, 2001).

### **Conclusiones sobre la prevalencia de los trastornos depresivos**

A modo de resumen, en la Tabla 1.6 se han recogido los resultados más relevantes relativos a la prevalencia de los trastornos depresivos y los trastornos del estado de ánimo que se han encontrado en los estudios epidemiológicos que se han revisado a lo largo de este epígrafe dedicado a la prevalencia de los trastornos depresivos. Se puede observar en la Tabla 1.6 que los datos sitúan la prevalencia en la población general del trastorno depresivo mayor a los 12 meses entre el 4% y el 10%, con la mayoría de los estudios cerca del 5.5%. En el caso de la prevalencia vital del trastorno depresivo mayor, los datos la sitúan entre el 10% y el 18.7%, siendo en el estudio ESEMeD para España del 10.6%.

De los datos de la Tabla 1.6, también se desprende que existen diferencias significativas en la prevalencia encontrada entre los distintos países (Narrow et al., 2002). Habitualmente es mayor esta diferencia cuanto mayores son las diferencias socio-culturales entre países, aunque puede ser que estas diferencias entren dentro del margen de error de la medición (Bromet et al., 2011). Además, estos autores consideran que:

“La amplia variabilidad de la estimación de la prevalencia vital de la depresión a los 12 meses es presumiblemente debida a la combinación de factores sustantivos (vulnerabilidad genética y factores de riesgo ambientales) y de medida (diferencias culturales en la aceptación y significado de los ítems, y las propiedades psicométricas de los instrumentos). Las diferencias en el diseño de los estudios también pueden estar relacionadas”. (Bromet et al., 2011, p. 2)

También se pueden observar diferencias entre estudios dentro de los mismos países. Además, se ha destacado que la prevalencia del trastorno depresivo mayor de los estudios que utilizan criterios DSM-IV y DSM-III-R son más altas que en aquellos en los que se usan criterios DSM-III (Hasin et al., 2005).

**Tabla 1.6.** Prevalencia vital y a los 12 meses del trastorno depresivo mayor (TDM) y de los trastornos del estado de ánimo (TEA) en los principales estudios epidemiológicos

<b>Autores</b>	<b>Estudio</b>	<b>Población (Nº países)</b>	<b>Instrumento de medida</b>	<b>Prevalencia 12 meses de TDM</b>	<b>Prevalencia vital de TDM</b>	<b>Prevalencia 12 meses de TEA</b>	<b>Prevalencia vital de TEA</b>
Bourdon et al., 1992	NCA	EE. UU.	DIS	5.9	6.3	6.3	8.3
Kessler et al., 1994	NCS	EE. UU.	CIDI	10.3	17.1	11.3	19.3
Kessler et al., 2003 <sup>c</sup>	NCS-R	EE. UU.	CIDI	6.8	16.9	9.7	21.4
Kessler et al., 2007 <sup>a</sup>	WMH	EE. UU.	CIDI 3.0			21.4	31.4
Hasin et al., 2005	NESARC	EE. UU.	AUDADIS-IV	5.28	13.23		
Bromet et al., 2011	WMH	Altos ingresos	CIDI 3.0	5.5	14.6		
Bromet et al., 2011	WMH	Bajos ingresos	CIDI 3.0	5.9	11.1		
Alonso et al., 2010	ESEMeD	Europa (6)	CIDI 3.0	3.9	12.8	4.2	14
Lépine et al., 1997	DEPRES	Europa (6)	MINI DEPRES	6.8 (6 meses)			
Ayuso-Mateos et al., 2001	ODIN	Europa (5)	SCAN			8.56 (1mes) <sup>a</sup>	
Wittchen y Jacobi, 2005	Metaanálisis	Europa(27)	Varios	8.3		9.1 <sup>a</sup>	
Jenkins et. al., 2003	NPMS	Inglaterra	CIS-R	2.3 (7 días) <sup>b</sup>			
Bijl et al., 1998	NEMESIS	Holanda	CIDI 1.1	5.8	15.4	7.6	19.0
de Graaf et al., 2011	NEMESIS 2	Holanda	CIDI 3.0	5.2	18.7	6.1	20.2
Jacobi et al., 2004	GHS-MHS	Alemania	CIDI	10.7 <sup>a</sup>	17.1 <sup>a</sup>	11.9	18.6
Haro et al., 2006a	ESEMeD	España	CIDI 3.0	4	10.6	4.3	11.4

<sup>a</sup> = Trastornos depresivos (excluye trastorno bipolar) durante el último mes; <sup>b</sup> = episodio depresivo; <sup>c</sup> = última actualización disponible en Harvard School of Medicine (2007). Fuente: Elaboración propia.

Respecto al trastorno distímico, en los estudios epidemiológicos desarrollados por la OMS en Europa, su prevalencia es aproximadamente un tercio de la prevalencia del trastorno depresivo mayor (p. ej., en España la prevalencia a los 12 meses es de 1.5%, siendo en el caso de la depresión mayor de 4.5%) (Kessler et al., 2007a; Haro et al., 2006a).

En conclusión, los datos actuales sobre la prevalencia del trastorno depresivo mayor y el trastorno distímico sugieren que los trastornos depresivos son uno de los principales problemas de salud a nivel mundial. Los estudios epidemiológicos de la depresión mayor han confirmado que es un trastorno extremadamente frecuente, “que está asociado con una variedad sorprendentemente extensa de morbilidad” Scott (2008, p. 2).

Como señalan Kessler et al. (2003, p. 3095):

“El trastorno depresivo mayor es un trastorno común, ampliamente distribuido en la población, y usualmente la gravedad en los síntomas está asociada al grado de disminución en el funcionamiento del sujeto. Aunque el incremento en los tratamientos es alentador, su inadecuación es una seria preocupación.”

Sin embargo, también existen voces críticas con estos estudios epidemiológicos (p. ej., Kira y Kutchins, 2002). Algunos autores consideran que los umbrales diagnósticos son demasiado bajos en estos estudios, indicando prevalencias demasiado altas para las actuales políticas de reparto de recursos (Regier y Narrow, 2002, en Mullins-Sweatt y Widiger, 2009). Estos autores opinan que sería imposible cubrir la atención a tan alto porcentaje de pacientes con trastornos, de modo que se corre el riesgo de que el diagnóstico deje de ser útil como posible indicador de la necesidad de tratamiento.

### 1.2.2. Características epidemiológicas asociadas de los trastornos depresivos

#### Edad de inicio

Según los principales estudios epidemiológicos realizados hasta la fecha, la **edad de inicio** de los trastornos depresivos se sitúa entre el inicio de la edad adulta y los 40 años (Hasin et al., 2005; Kessler et al., 2007; Kessler et al., 2010). En la encuesta WMH de la OMS se encontró que la media de edad de inicio de los trastornos del estado de ánimo se situaba en el rango de finales de los 20 y los 40 años (29-43) (Kessler et al., 2007), de manera que la prevalencia es baja hasta la adolescencia, aumentando de manera rápida, para tener un incremento más gradual después de los cuarenta (Kessler et al., 2007). En esta misma línea, el estudio NESARC encontró una media de edad de inicio de 30.4 años (Hasin, et al. 2005). La aparición por primera vez de trastornos depresivos en la tercera edad es menos prevalente que en edades más tempranas (Kessler et al., 2003; Kessler et al., 2010). Los autores añaden que la distribución de la edad de inicio es parecida a la de algunos trastornos de ansiedad como la ansiedad generalizada, el pánico o el estrés postraumático.

En España, según los estudios incluidos dentro del proyecto ESEMeD, la edad de inicio de los trastornos depresivos se sitúa a partir de los 40 años (Haro et al, 2006a). En cualquier caso, todos estos estudios son retrospectivos, por lo que se ha señalado la posibilidad de que la edad de inicio detectada mediante las entrevistas diagnósticas esté siendo distorsionada por los sesgos en la recuperación de la información por parte de los sujetos (Kessler et al., 2010). Si se le pregunta a una persona por sus episodios depresivos anteriores, evaluar si cumplía o no los criterios diagnósticos décadas antes es al menos problemático.

La guía NICE (2010) advierte de la importancia de la detección temprana de la sintomatología depresiva debido a que si la edad de inicio es menor de los 20 años, la probabilidad de recaída y cronicidad es mayor.

### Diferencias de género

Como se ha comentado antes, los trastornos depresivos son muy frecuentes en la población general, pero en el caso de las mujeres se encuentra de forma consistente que su frecuencia es aún mayor (véase la Tabla 1.7). Diversos autores han calculado el tamaño de estas diferencias entre géneros: 2 veces (Nolen-Hoeksema y Hilt, 2009) y entre 1.5 y 2.5 veces más altas en mujeres que en hombres (Waraich et al., 2004, en NICE, 2010). Estas tasas son estables en el rango de edad entre los 18 y los 64 años.

**Tabla 1.7.** Prevalencia a los 12 meses de los trastornos depresivos en hombre y mujeres en estudios europeos de prevalencia

<b>Autores</b>	<b>Estudio</b>	<b>Población (Nº Países)</b>	<b>Categoría Diagnóstica</b>	<b>Mujeres</b>	<b>Hombres</b>
Alonso et al., 2004a	ESEMeD	Europa (6)	Trastorno depresivo mayor	16.5%	8.9%
Haro et al., 2006a	ESEMeD	España	Trastornos depresivos	15.8%	6.7%
Wittchen y Jacobi, 2005	Meta- análisis	Europa (27)	Trastornos depresivos	9.1%	6.1%



### **¿Está creciendo con el tiempo la prevalencia de los trastornos del estado de ánimo?**

Kessler, (2007, p. 13) afirma que “En encuestas anteriores los análisis retrospectivos sobre la edad de inicio han mostrado patrones consistentes que sugieren que la prevalencia vital de los trastornos de ansiedad y del estado de ánimo se han incrementado”, aunque añade que no se ha informado de patrones similares en los estudios de EDEMeD.

En este sentido, se ha afirmado que los trastornos del estado de ánimo están aumentando en los países industrializados desde 1910 (Cross-National Collaborative Group, 1992; Klerman, 1988; Klerman et al., 1985, en APA, 2002). Los aumentos detectados por algunos estudios pueden deberse a que la generación nacida tras la Segunda Guerra Mundial tenga una prevalencia de trastornos mentales más alta que las generaciones anteriores, pero también más elevada que los jóvenes actuales (Hasin, et al. 2005).

Por el contrario, no se han encontrado diferencias generacionales significativas en estudios con grandes muestras como el NCS (Kessler, 1994) y el NCS-R (Kessler et al., 2003), y en las encuestas nacionales de Inglaterra (Jonas et al., 2011) y de Holanda (de Graaf et al., 2011) las diferencias no son significativas. Será necesario que las encuestas se sigan realizando durante un periodo más prolongado con metodología comparable a la actual para tener información más precisa.

### **Distribución de la gravedad de la sintomatología de los trastornos depresivos**

Es importante desde el punto de vista sociosanitario conocer cuál es la distribución de la gravedad de la sintomatología de los sujetos con trastorno depresivo mayor, puesto que a mayor gravedad del trastorno más deterioro se

produce en la persona. En el estudio WMH de la OMS se incluyeron dentro de la entrevista CIDI 3.0 las preguntas estructuradas necesarias para la valoración de la gravedad de la sintomatología. De manera consistente se encontró que los trastornos depresivos leves fueron los más frecuentes, seguidos de los moderados y graves. En el caso de España se puede observar en la Tabla 1.8 que se sigue esta misma distribución, con cifras muy similares a la de otros países de nuestro entorno como Italia, y, de hecho, en España, solamente el 1% de la muestra fue diagnosticada con trastorno depresivo mayor grave (Demyttenaere et al., 2004).

**Tabla 1.8.** Prevalencia a los 12 meses del trastorno depresivo mayor según la entrevista CIDI en función de las distintas categorías de gravedad del trastorno y de los diversos países que participaron en la Iniciativa *World Mental Health* de la Organización Mundial de la Salud

Países	Grave	Moderado	Leve	Total
Colombia	5.2	6.6	5.9	17.8
México	3.7	3.6	4.9	12.2
EE.UU.	7.7	9.4	9.2	26.3
Belgica	2.4	3.3	6.4	12.0
Francia	2.7	6.1	9.7	18.4
Alemania	1.2	3.3	4.5	9.1
Italia	1.0	2.9	4.3	8.2
Holanda	2.3	3.7	8.8	14.8
España	1.0	2.9	5.3	9.2
Ucrania	4.8	7.4	8.2	20.4
Líbano	4.6	6.2	6.1	16.9
Nigeria	0.4	0.5	3.8	4.7
Japón	1.5	4.1	3.2	8.8

No se pueden realizar estimaciones sobre la gravedad de los episodios depresivos ocurridos más allá de los 12 meses debido a que la probabilidad de existencia de sesgos de medida es muy significativa.

En el estudio NCS-R, utilizando una metodología distinta, se calculó mediante el instrumento QIDS-SR la gravedad de los síntomas de los trastornos depresivos, y se encontró una distribución distinta a la encontrada en la mayoría de los países participantes en la iniciativa WMH de la OMS. En concreto, se encontró que el 10.4% del total de los trastornos depresivos eran leves, 38.6% moderados, 38.0% graves y 12.9% muy graves (Kessler et al., 2003). Agrupando las categorías para comparar estos resultados con los de la iniciativa WMH de la OMS, parece que en los Estados Unidos de América el 10,4% de los trastornos depresivos son leves, el 38,6% moderados y el 59% graves-muy graves, lo cual contrasta con la distribución de la gravedad en los datos ya comentados de la iniciativa WMH de la OMS. Si estas diferencias tan importantes entre ambos estudios se deben a diferencias metodológicas, a diferencias culturales o a ambas, no es posible de determinar actualmente, aunque, dado el amplio número de países que fueron evaluados en el estudio WMH de la OMS, una mayor prevalencia de las depresiones leves y moderadas frente a las graves parece que es más plausible o al menos más generalizable internacionalmente.

### **Comorbilidad**

La comorbilidad se define como la presencia de un trastorno simultáneo o sucesivo con otro en un periodo temporal determinado (Bijl, et al., 1998). La comorbilidad ha sido considerada como la norma, más que como la excepción, en los adultos con trastornos depresivos, especialmente con los trastornos de ansiedad (Goodwin, et al., 2006). Se considera que en torno al 50% de los sujetos diagnosticados con un trastorno mental, cumple además criterios para otro

diagnóstico (Demyttenaere et al., 2004 y Kessler, Chiu, Demler y Walters, 2005). Además, “la comorbilidad está relacionada con un incremento en la gravedad, una mayor duración del trastorno, un nivel mayor de interferencia y un incremento en el uso de los servicios sanitarios” (Alonso et al., 2004a, p. 29).

A través de un análisis de los datos obtenidos del proyecto ESEMeD para los 6 países europeos analizados, se encontró una alta asociación entre los trastornos de estado de ánimo y los trastornos de ansiedad. De las personas con diagnóstico de trastorno del estado de ánimo, el 41.7% tenía además un trastorno de ansiedad o un trastorno relacionado con el consumo de sustancias, siendo la categoría diagnóstica con mayor porcentaje de comorbilidad. Dentro de los trastornos del estado de ánimo, la distimia tenía un mayor porcentaje de comorbilidad (73.3%) en comparación con la depresión mayor (53.1%). No parece que existieran diferencias significativas entre países. Además, la mayor comorbilidad entre trastornos específicos se dio entre la depresión mayor y el trastorno de ansiedad generalizada (33.7%) y en segundo lugar el trastorno de pánico (29.4%). En el caso de la distimia, la comorbilidad se produjo con más frecuencia con el trastorno de agorafobia, con el 29.4% (Alonso et al., 2004a).

En el estudio español, más de la mitad de los pacientes diagnosticados con trastorno depresivo tenía además un segundo diagnóstico asignado: el 55.8% de los pacientes diagnosticados de depresión mayor cumplían además los criterios de algún otro trastorno mental, un 42.2% un trastorno de ansiedad y un 27.1% distimia (Haro, 2006a).

Los resultados de la encuesta NCS-R señalan que más de un 40% de los sujetos que cumplían los criterios para un diagnóstico de trastorno mental cumplían también los criterios de otro diagnóstico (Kessler, et al., 2005).

Sin embargo, se ha comentado que el hecho de que la evaluación sea realizada a través de entrevistas estructuradas en lugar de entrevistas clínicas, puede haber inflado artificialmente la estimación de la comorbilidad entre el diagnóstico de trastorno depresivo mayor y otros trastornos debido a la superposición de los síntomas principales (Kessler et al., 2010).

### **Suicidio y depresión**

La Organización Mundial de la Salud (2000) estima que el número de personas propensas a realizar conductas suicidas en el mundo es de un millón al año y es una de las 3 principales causas de muerte entre los 15 y los 35 años de edad. Los trastornos mentales son un importante factor de riesgo para las conductas suicidas, sobre todo en países desarrollados (Hawton y van Heeringen, 2009). En concreto, las personas con un trastorno depresivo mayor tienen 4 veces más probabilidades de suicidarse (Sartorius, 2001, en NICE, 2010), mientras que en los casos más graves el riesgo se multiplica por 20. Bostwick y Pankratz, (2000, p. 20) señalan que “no hay factores de riesgos, incluidos los subtipos de clasificación diagnóstica, que se hayan mostrado fiables en la predicción del suicidio”.

En una revisión se encontró que la mitad de las personas que llegan a consumir el suicidio tenían un trastorno depresivo (Cavanagh, Carson, Sharpe y Lawrie, 2003). Además cerca del 4% de los individuos deprimidos fallecen mediante el suicidio, pero el riesgo es mayor en hombres que en mujeres (Coryell y Young, 2005). Se ha señalado que es más probable que las conductas suicidas se lleven a cabo durante el primer episodio depresivo y cuando están asociadas a trastornos de personalidad y a un perfil impulsivo (Hawton y van Heeringen, 2009).

### 1.3.3. Importancia sociosanitaria de los trastornos depresivos

Los problemas sociales y económicos derivados de los trastornos mentales son mayores de lo esperado (Knapp, 2006). Con el desarrollo de la epidemiología y la formulación de métodos de valoración del coste de las enfermedades, se han realizado durante las últimas décadas estudios globales dirigidos a conocer el peso de las patologías sobre la sociedad. El informe *Global Burden of Disease* (Murray y López, 1997) señaló que dentro de las enfermedades de la población general, los trastornos mentales ocupan un espacio muy significativo como problema sociosanitario. El trastorno depresivo mayor es uno de los trastornos más relevantes desde este punto de vista debido a que tiene una alta prevalencia, una edad de inicio temprana, un curso recurrente y además está asociado a un gran coste personal y social (Alonso et al., 2004a; Gabilondo et al., 2010; Kessler, 2007).

La depresión mayor fue detectada como la cuarta causa en el mundo de años de vida productiva perdidos por discapacidad o DALYs (*Disability Adjusted Life Years*), después de las infecciones respiratorias, las diarreas y los problemas perinatales (Murray y López, 1997). Üstün et al (2004) detectaron que la depresión era en el año 2000 la principal causa de DALYs en todo el mundo. Estudios posteriores señalaron que en los países con nivel medio o bajo de ingresos, la depresión unipolar era la 7ª causa de DALYs, mientras que en los países con altos ingresos, la depresión ocupaba el tercer lugar (López, Mathers, Ezzati, Jamison y Murray, 2006).

En la misma dirección, en el informe publicado por la OMS acerca de las enfermedades más importantes y las proyecciones desde el año 2004 al 2030 (OMS, 2004), se indica que la depresión unipolar es la principal causa de años vividos con discapacidad o YLDs (*Years Lived with Disability*), tanto en países altos

ingresos como en países con niveles de ingresos bajo o medios. Ese mismo año se situaba a la depresión unipolar como tercera causa de DAYLs, por detrás de las infecciones respiratorias y la diarrea, pero, sin embargo, se proyectaba para el año 2030 que la depresión unipolar será la principal casusa de DAYLs a nivel mundial (OMS, 2004).

Wittchen y Jacobi (2005), en una revisión sobre el coste de la depresión en Europa, destacan que los trastornos depresivos junto con el trastorno de pánico, son los trastornos mentales que más discapacidad producen, no solamente por encima de otros trastornos psiquiátricos, sino también de trastornos somáticos como la diabetes o las enfermedades cardíacas.

Aunque la mayoría de las personas que tienen trastornos mentales no buscan ayuda (Goldberg y Huxley, 1980), la depresión se asocia a una mayor utilización de servicios sanitarios, a un mayor deterioro de la actividad física, de las relaciones sociales y del rendimiento laboral, y a una percepción más pobre de la salud actual (Wang, Simon y Kessler, 2003). Además, existe una relación clara y directa entre la gravedad del trastorno y el grado de deterioro (NICE, 2010).

En función de los resultados encontrados en el proyecto ESEMeD, Alonso et al. (2004b, p. 44) señalan que: “En conjunto, el impacto de los trastornos mentales en la discapacidad y la calidad de vida parece similar o superior al impacto de los trastornos físicos crónicos más comunes” Alonso et al. (2004b, p. 44). La fobia social, el trastorno de pánico, el trastorno de estrés postraumático y la depresión mayor suponen un “impacto similar y relativamente alto” en la pérdida de calidad de vida y una mayor pérdida de días de trabajo. Estos autores nombran dos limitaciones importantes de este estudio, que se pueden generalizar a la mayoría de los estudios sobre el coste de la depresión: los sujetos con trastornos afectivos pueden tener un sesgo pesimista sobre su estado de salud o su grado de deterioro

y además los estudios no establecen si las dificultades de funcionamiento o el deterioro físico son la causa o la consecuencia del trastorno depresivo.

### **Los costes económicos de la depresión**

Los altos costes económicos de la depresión están relacionados con dos aspectos diferenciados: los gastos en cuidados sociosanitarios y los días de trabajo perdidos. Según Paykel (2006) las estimaciones que se han realizado en Estados Unidos (población de 270 millones de personas) sobre el coste total de la depresión varían entre los 16 000 millones de dólares (Studemire et al., 1986), 30 000 millones de dólares (Rice y Miller, 1995) y 43 000 millones de dólares (Greenberg et al., 1993).

Los estudios europeos varían significativamente unos de otros (Lothgren, 2004; Sobocki y Wittchen, 2005). En una revisión realizada por Sobocki, Jönsson y Rehnberg (2006) sobre los costes económicos de la depresión en Europa, se analizaron 28 países, con una población total de 466 millones de personas. El coste estimado en el año 2004 fue de 118 000 millones de euros (42 000 millones en costes directos y 76 000 millones en costes indirectos). “Es por tanto el trastorno mental que más gasto provoca, con un 33% del total, correspondiendo con un 1% de la economía europea” (Sobocki, Jönsson y Rehnberg, 2006, p. 87). Estos autores cifran el coste económico en España de la depresión en 5 005 millones de euros.

En un estudio publicado en 2009 sobre los costes directos e indirectos derivados de los trastornos mentales en el año 2002 en España, se calcularon los días perdidos de empleo, los costes sanitarios formales derivados de ingresos hospitalarios y de la atención ambulatoria, la medicación, las muertes prematuras, así como los costes derivados de la atención informal a los pacientes. El coste total



se calculó en 7 019 millones de euros, siendo un 39% de esa cantidad debido a costes directos y un 42% debido a pérdidas de productividad laboral (el 0.98% del producto interior bruto) (Oliva-Moreno, López-Bastida, Montejo-González, Osuna-Guerrero y Duque-González, 2009).

Sin embargo en otro estudio se señala que el coste total de la depresión en España sería 745 millones de euros al año, 535 millones de los cuales estarían asociados a gastos directos del manejo y tratamiento del paciente, y el resto sería atribuible a la pérdida de productividad laboral (OFisalud en Ayuso et al., 2008).

# **Capítulo 2. EL DIAGNÓSTICO DE LOS TRASTORNOS DEPRESIVOS**



**“Antes de contestar la pregunta de ¿Cómo medimos algo?, debemos contestar primero a la pregunta: ¿qué es exactamente ese algo?”**

**Joiner, Walker, Pettit, Pérez y Cukrowicz, 2005, p. 267**

Este capítulo está dedicado a la naturaleza del diagnóstico de los trastornos mentales, específicamente de los trastornos depresivos, así como a su utilidad y a las dificultades existentes para su realización.

## **2.1. El papel del diagnóstico en psicología clínica**

El debate sobre la utilidad de las categorías diagnósticas ha ocupado siempre un lugar central en el campo de la salud mental y más concretamente en el campo de la psicología clínica. Durante la primera mitad del siglo XX la escuela psicodinámica desarrolló un sistema de clasificación de patologías basado en su propia teoría. Una parte de los profesionales de la psicología clínica (entre ellos los psicólogos conductuales) rechazó el uso de clasificaciones psicopatológicas, aunque en cierta manera continúan utilizando etiquetas que describen los procesos de aprendizaje en los pacientes. Desde la aparición de las categorías descriptivas desarrolladas a finales de los años 70 del siglo XX, una parte importante de los investigadores y psicólogos clínicos utilizan de manera habitual estas categorías diagnósticas. A continuación se describen las principales características de las clasificaciones diagnósticas más ampliamente utilizadas en la actualidad, sus ventajas y sus desventajas.

### **2.1.1. Definición de trastorno mental**

El manual diagnóstico DSM-IV-TR (APA, 2002) describe el concepto de trastorno mental como:

“Un síndrome o un patrón comportamental o psicológico de significación clínica, que aparece asociado a un malestar (p. ej., dolor), a una discapacidad (p. ej., deterioro en una o más áreas de funcionamiento) o a un riesgo significativamente aumentado de morir o de sufrir dolor, discapacidad o pérdida de libertad. Además, este síndrome o patrón no debe ser meramente una respuesta culturalmente aceptada a un acontecimiento particular (p. ej., la muerte de un ser querido). Cualquiera que sea su causa, debe considerarse como la manifestación individual de una disfunción comportamental, psicológica o biológica” (APA, 2002, p. XXI).

El propio manual asume que el concepto de trastorno mental no permite una descripción operacional “que especifique adecuadamente los límites del concepto” (APA, 2002, p. XXI), de forma que aunque todos los indicadores o criterios que incluye la anterior definición de trastorno mental son útiles para definir el concepto y entender algún tipo de trastorno mental, ninguno de esos indicadores o criterios equivale al concepto y cada tipo de trastorno requiere una definición. Es decir, es importante señalar que cada categoría asume su propia definición de trastorno (p. ej., malestar, descontrol, limitación, incapacidad, inflexibilidad, irracionalidad, patrón sindrómico, etiología y desviación estadística). En el caso de la depresión, quizá los componentes clave que describen mejor su naturaleza sean el *patrón sindrómico de síntomas distintivos, especialmente estado de ánimo depresivo y anhedonia, que provoca malestar e incapacidad en el paciente*.

Actualmente, los dos principales sistemas de clasificación utilizados en el diagnóstico de los trastornos mentales son el del “Manual diagnóstico y estadísticos de los trastornos mentales” (DSM) editado por la Asociación Americana de Psiquiatría (APA, 2002) y el de la “Clasificación Internacional de Enfermedades”

(CIE) editado por la Organización Mundial de la Salud (OMS, 1993). Ambos sistemas son categoriales y multiaxiales e intentan incluir criterios diagnósticos objetivables. Son además sistemas politéticos, puesto que para se pueda asignar un diagnóstico es suficiente cumplir con un cierto número de criterios sintomáticos o diagnósticos de entre una lista de criterios propuestos, sin que sea necesario que se cumplan todos los criterios de la lista, aunque, por otro lado, puede también haber criterios necesarios para hacer un determinado diagnóstico.

Las etiquetas diagnósticas, como la de trastorno depresivo mayor, permiten conocer propiedades generales sobre estas categorías, pero los trastornos mentales no son grupos naturales formados por elementos homogéneos (Kendler y Zachar, 2008). A pesar de proporcionar información útil, no son una realidad diferenciable con límites claros. Las clasificaciones diagnósticas en psicología son construcciones teóricas que intentan dotar de un cierto orden a la gran variedad del comportamiento humano, “no son hechos, sino fundamentalmente nombres que utilizamos para organizar la realidad” (Vázquez, Sánchez y Romero, 2011, p. 92). Por su parte, el diagnóstico es “el proceso por el cual se determina si los problemas que afectan a una persona cumplen todos los criterios de un trastorno psicológico, que vienen especificados en las clasificaciones diagnósticas” (Muñoz, 2003, p. 60).

El estado de confusión teórica previo al DSM-III (APA, 1983) tenía su reflejo en las dos primeras versiones del DSM, DSM-I (APA, 1952) y DSM-II (APA, 1968). Los sistemas de clasificación de estos dos manuales eran caóticos, imprecisos y elaborados mediante definiciones narrativas, lo que provocó que hasta los años 70 del siglo XX el diagnóstico de los trastornos mentales (también en el caso de los trastornos depresivos) fuera poco fiable y dependiente en exceso de las propias teorías de cada psicólogo. Durante los años 70 del siglo pasado se produjo en el estudio de la salud mental una auténtica revolución en los sistemas de clasificación

que los situaron en un lugar destacado como herramientas fundamentales para los profesionales de la salud mental. Así, por ejemplo, el aparente ateoricismo de los manuales diagnósticos DSM y CIE actuales y la naturaleza descriptiva de sus criterios sintomáticos, permite a los especialistas en salud mental de distintas corrientes comprobar si sus intervenciones son eficaces en la medida en la que los pacientes dejan de cumplir los criterios diagnósticos de los trastornos.

La corriente neokraepeliana de la década de los 70 del siglo XX señalaba que debía recuperarse una clasificación más descriptiva de la sintomatología. La sintomatología descrita en el DSM-III (APA, 1983), elaborada a través de un proceso de consenso entre expertos de diversas corrientes, facilitó la adopción de un lenguaje común que permitía la comunicación entre especialistas, aunque en ocasiones se llegó a acuerdos que no tenían en cuenta solo cuestiones científicas de peso. Estos cambios produjeron resultados: mayor fiabilidad, mayor concreción, ejes diagnósticos para recoger información suplementaria, estudios de campo previos y mayor cercanía a problemas conductuales, especialmente en la infancia (Blashfield y Livesley 2008). El éxito de los criterios diagnósticos del DSM-III y de sus versiones sucesivas, así como el éxito de los criterios de la CIE-10, han cambiado el modo en el que los clínicos realizan las evaluaciones y los tratamientos de los trastornos mentales.

### **2.1.2. Utilidad del diagnóstico de los trastornos mentales**

Los sistemas de categorización “proporcionan un ahorro en la memoria (o una reducción en la carga cognitiva); permiten la manipulación de los objetos mediante la simplificación de las relaciones entre ellos; y permiten generar hipótesis y predicciones” (Jablensky y Kendell, 2002, p. 3). De hecho, “la actividad clasificatoria es una expresión natural del funcionamiento cognitivo humano” e incluso aquellas teorías psicológicas que se oponen a las clasificaciones de manera

global, suelen incluir categorías explicativas de manera quizá más informal y con validez no contrastada (Vázquez, 1990, p. 656). Por ejemplo, desde las teorías conductuales se agrupan procesos y síntomas que etiquetan los problemas de los pacientes. Por tanto clasificar es un “hecho necesario y fundamental” (Kendell, 1975, en Blashfield y Livesley, 2008). Sin embargo, se ha de entender que al contrario de otros procesos clasificatorios del hombre, las categorías referidas a la salud mental no señalan con frecuencia fronteras claras y nítidas entre quienes presentan o no presentan un trastorno, como buscaban las teorías neokraepelianas. Además, mientras no se encuentren estas fronteras y según se acumulen conocimientos sobre cada categoría, estas clasificaciones son criticables y necesitan una continua revisión.

En este sentido, el propio manual DSM-IV-TR de la APA (2002, p. XV) señalaba que:

“La utilidad y credibilidad del DSM-IV exigen que se centre en objetivos clínicos, de investigación y educacionales, y se apoye en fundamentos empíricos sólidos. Nuestra prioridad ha sido la de proporcionar una guía útil para la práctica clínica. Mediante la brevedad y concisión en los criterios, la claridad de expresión y la manifestación explícita de las hipótesis contenidas en los criterios diagnósticos, esperamos que este manual sea práctico y de utilidad para los clínicos. Otro objetivo de este documento, aunque adicional, es el de facilitar la investigación y mejorar la comunicación entre los clínicos y los investigadores. Hemos intentado asimismo que sirva de herramienta para la enseñanza de la psicopatología y para mejorar la recogida de datos en la información clínica hasta ahora vigente”.



A continuación se resumen las principales **ventajas del uso de los sistemas clasificatorios actuales**, válidas también para las categorías diagnósticas de los trastornos depresivos:

- Uno de los principales objetivos logrados por los sistemas de clasificación es el de **mejorar la comunicación** con otros profesionales mediante la unificación de un lenguaje común (p. ej., APA, 2002; Jablensky y Kendell 2002). Aunque los clínicos puedan tener modelos explicativos diferentes entre sí sobre lo que le sucede al paciente, poder utilizar etiquetas que resumen la información sintomática de manera descriptiva supone un avance fundamental en el campo de la salud mental para la comunicación entre profesionales.
- Del mismo modo, gracias al consenso alcanzado durante las últimas décadas en torno a los sistemas de clasificación, tanto la población general como los propios pacientes se han **familiarizado con una serie de términos o etiquetas diagnósticas** que tratan de explicar y desculpabilizar ciertos problemas. Además, el uso de este tipo de términos diagnósticos puede hacer que más personas rompan el estigma relacionado con la salud mental y soliciten atención especializada, ya que supondrían que consideran sus problemas como cualquier otro tipo de problema de salud, normalizando su existencia y la búsqueda de ayuda profesional al respecto (Üstün, Chatterji y Andrews, 2002). No obstante, tratando de impedir la asimilación de la etiqueta diagnóstica a la persona, el DSM-IV-TR (APA, 2002) destaca la necesidad de utilizar expresiones como “un individuo con depresión” en lugar de “un depresivo”, precisamente para evitar el error de clasificar personas en lugar de trastornos.
- Durante las evaluaciones clínicas las categorías diagnósticas permiten **organizar la recogida de información sobre el paciente**. En este sentido,

ayudan a la identificación de las conductas problema y aportan información sobre la correlación existente entre determinados síntomas, de manera que se dé coherencia a la sintomatología de los sujetos (Muñoz, 2003; Vizcarro, 2000). Esta mejora ha permitido introducir al personal sanitario no especializado en la evaluación e intervención en salud mental (Üstün et al., 2002). De cualquier modo, una evaluación psicológica integral obviamente trasciende el diagnóstico categorial psiquiátrico.

- Los sistemas de clasificación han aportado información sobre el **curso esperable de los trastornos** de un paciente (Üstün et al., 2002; Vázquez, et al., 2011). Por ejemplo, en el caso de la naturaleza episódica de la depresión mayor, se ha llamado especialmente la atención sobre la alta frecuencia de recaídas durante los primeros meses tras el tratamiento. Este hecho ha guiado a los clínicos a perseguir la eliminación de la sintomatología residual como una estrategia eficaz para la prevención de recaídas.
- Los sistemas de clasificación también han permitido la posibilidad de estudiar la **eficacia diferencial de distintos tipos de tratamiento** (Pérez, Fernández, Fernández y Amigo, 2005; Üstün et al., 2002). En este sentido existe una gran discusión sobre la importancia que se debe otorgar al diagnóstico en la toma de decisiones del tratamiento. Por un lado se ha defendido la idea de que las etiquetas diagnósticas recogen información demasiado heterogénea como para asumir que todos los pacientes con dicha etiqueta deben ser tratados del mismo modo, por lo que el diagnóstico no indicaría directamente qué intervención se debe realizar para una persona en concreto. Por el contrario, pensar que no se deben manualizar los tratamientos en función del diagnóstico sería perder información útil (Norcross, 2002, en Pérez, Fernández, Fernández y Amigo, 2005).

- El diagnóstico y los sistemas de clasificación son también de utilidad para mejorar el **reparto de los recursos sociosanitarios**, puesto que permite realizar, por ejemplo, estudios epidemiológicos y de costes-beneficios que permitan conocer qué trastornos son los más frecuentes, cuáles tienen consecuencias más graves para el individuo y la comunidad, cuáles tienen mayor probabilidad de remitir o recaer, etc., y, en función de esta información, dar prioridad y repartir los recursos preventivos y terapéuticos asignados a la salud mental.
- Otra ventaja de los sistemas de clasificación tiene que ver con la sistematización de la **enseñanza a los estudiantes de salud mental** (Chatterji y Andrews, 2002). En la actualidad, resulta difícil pensar en la investigación o en la docencia sin el uso de etiquetas, que sirven para estructurar los estudios, simplificar la formación y concretar procedimientos de evaluación más uniformes. A su vez, esta ventaja no está exenta de riesgos, ya que se ha señalado que el excesivo énfasis en los sistemas de clasificación puede llevar a alumnos y profesores a utilizar los manuales de diagnóstico como si fuesen libros de texto de psiquiatría y psicología clínica, “lo cual conduce a que los criterios de diagnóstico se consoliden como *principios sagrados*” (Alarcón, 2010, p. 38).
- Las categorías diagnósticas también **ayudan a organizar y dirigir la investigación**. En este sentido, las actuales clasificaciones han ayudado a acercarse a **las causas de los trastornos** (o a sus factores de riesgo). Los clínicos deben ser conscientes de que las categorías diagnósticas son conceptos, justificados porque aportan organización y explicación de experiencias clínicas complejas y permiten realizar predicciones sobre el pronóstico y guiar el proceso de toma de decisiones para el tratamiento

(Blashfield y Livesley, 2008), pero, sin embargo, las bases etiológicas de dichas categorías diagnósticas son multicausales, complejas, no uniformes y en ocasiones compartidas para varios diagnósticos. Es cierto, además, que los manuales diagnósticos no pueden ser ateóricos por completo, y sería ingenuo pensar que los actuales lo son (Millon, 1991, en Jablensky y Kendell, 2002). Entender la teoría que aglutina los actuales sistemas clasificatorios es fundamental para su correcta utilización.

Estas ventajas, junto con circunstancias sociales y económicas, han provocado la expansión del uso de los sistemas de clasificación y de los manuales diagnósticos. No obstante, existen también **limitaciones** en torno a su uso:

- Un aspecto negativo de los sistemas diagnósticos es el **estigma social y el estigma que se impone el propio individuo** al recibir un diagnóstico (Corrigan, 2005, y Goffman, 1963, en Vázquez et al., 2011). En ocasiones, este estigma puede llegar a producir más problemas que el propio trastorno (Hinshaw y Stier, 2008, en Mullins-Sweatt y Widiger, 2009). Cuando los clínicos o los propios pacientes entienden de manera no adecuada las etiquetas diagnósticas, éstos últimos pueden adoptar un rol de enfermo que marca de manera inevitable su conducta, provocando indefensión, baja autoestima o aislamiento social.
- Desde las diversas corrientes psicológicas se ha señalado que existen procesos psicológicos y sociales que aportan información más útil para la evaluación y tratamiento de los problemas de los sujetos y que no quedan recogidos ni reflejados en las etiquetas diagnósticas. En este sentido, aunque la psicología en general ha ido adaptando su lenguaje (así como la medición de la eficacia de los tratamientos) a las clasificaciones diagnósticas, se considera que lo fundamental es detectar las causas y los mantenedores de la

sintomatología en el paciente y modificar los modos de afrontamiento disfuncionales.

- Algunos autores defienden la idea de que se debe evitar “la **tendencia a considerar como médicos o patológicos aquellos problemas que se encuentran dentro del espectro normal**” (Pincus, McQueen y Elinson, 2005, p. 139). Por ejemplo, se ha señalado que un diagnóstico no equivale a la necesidad de que una persona reciba medicación (Zisook y Kendler, 2007). Algunos especialistas en salud mental y algunos sectores sociales opinan que es un error que las categorías diagnósticas incluyan situaciones que difícilmente pueden considerarse como patológicas, en la medida en la que suponen reacciones esperables ante acontecimientos vitales, casi podría decirse que necesarias desde el punto de vista social. Tal es el caso, por ejemplo, de las manifestaciones de tristeza intensa ante la muerte de una persona cercana. En este sentido, Wakefield, Schmitz, First y Horwitz (2007) señalan la dificultad para establecer un criterio fiable y válido respecto al diagnóstico de trastorno depresivo cuando la sintomatología está relacionada con la reacción de una persona ante un suceso vital estresante (ya sea un duelo o cualquier otro acontecimiento significativo). Hay que recordar que, tal y como se ha mencionado antes, la definición del DSM-IV (APA, 2002) de los trastornos mentales ponía precisamente la reacción a la muerte de un ser querido como ejemplo de criterio de exclusión, es decir, como ejemplo de una respuesta culturalmente aceptada a un acontecimiento particular que no debería considerarse como trastorno mental. De manera provisional, en el desarrollo del DSM-5 se ha incluido la siguiente nota para ayudar al diagnóstico diferencial entre el duelo y el episodio depresivo mayor:

“La respuesta normal y esperada ante un evento relacionado con una pérdida significativa (por ejemplo, el duelo, la ruina financiera, un desastre natural), incluyendo sentimientos de tristeza intensa, rumiaciones sobre la pérdida, insomnio, bajo apetito y pérdida de peso, se puede parecer a un episodio depresivo. La presencia de síntomas como los sentimientos de desesperanza, las ideas suicidas (distinto del deseo de unirse con el ser querido que se ha perdido), el retardo psicomotor y el deterioro grave del funcionamiento general sugieren la presencia de un trastorno depresivo mayor adicional a la respuesta normal ante una pérdida significativa” (APA, 2012).

- Durante las últimas décadas se han producido mejoras en **la eficacia en la atención a los pacientes**, pero todavía es objeto de discusión hasta qué punto se deben atribuir estas mejoras al uso de las clasificaciones diagnósticas (NICE, 2010). Desde sus impulsores, existe la conciencia de que son necesarios estudios empíricos para determinar la utilidad de los diagnósticos y de sus sucesivos cambios (First et al., 2004). En este sentido, se ha sugerido que los beneficios obtenidos en la atención a los trastornos mentales no se pueden atribuir en su totalidad a la existencia de tratamientos específicos para cada categoría diagnóstica; otros factores inespecíficos y comunes a los tratamientos para todo tipo de trastornos como, por ejemplo, el reconocimiento de la angustia, la reinterpretación de los síntomas y el brindar esperanza y apoyo social contribuirían a obtener mejores resultados con los pacientes, con independencia de la etiqueta diagnóstica (Ormel et al., 1990, en NICE, 2010).
- La formulación multiaxial de los sistemas de clasificación permite recoger información sobre el **funcionamiento psicosocial** del paciente y su grado de

**malestar**, pero se ha señalado que esta información es poco detallada y se le otorga poca relevancia en el diagnóstico (Muñoz, 2003).

- Los clínicos no especializados en investigación tienen con frecuencia **dificultades para recordar y utilizar correctamente los criterios diagnósticos** de los sistemas de clasificación. Sin embargo, también parece cierto que las categorías diagnósticas más estables durante las últimas décadas tienen un alto grado de conocimiento y de aceptación por parte de los profesionales en salud mental (como es el caso del trastorno depresivo mayor) (Mullins-Sweatt y Widiger, 2009).
- La **asignación de una etiqueta diagnóstica a un individuo es un proceso problemático desde el punto de vista de su validez**. En contra de lo esperado por las teorías neokraepelianas, que suponen que las diferentes categorías diagnósticas representan “enfermedades” diferentes, es decir, alteraciones del funcionamiento biológico normal en cualquiera de sus niveles (p. ej., cromosómico, histológico, anatómico, fisiológico) que podrían detectarse mediante las correspondientes pruebas biológicas, no se han hallado pruebas o marcadores genéticos o neurobiológicos válidos y fiables para la detección de la mayoría de los trastornos mentales, incluyendo los trastornos depresivos.

### 2.1.3. Fiabilidad del diagnóstico de los trastornos mentales

Como se ha comentado anteriormente, los sistemas de clasificación diagnóstica mejoraron su fiabilidad a finales de los años 70 y principios de los 80 del siglo pasado. En este sentido, la operativización de los criterios diagnósticos permitió a los distintos especialistas fijar su atención en los mismos elementos durante la evaluación diagnóstica, lo que, a su vez, permitió mejorar tanto la fiabilidad interjueces como la fiabilidad test-retest (p. ej., Smith y Combs, 2010). No obstante, existen también autores que ponen en duda que las mejoras en fiabilidad

fueran realmente tan significativas: “fueron las declaraciones de su éxito, de cualquier modo, lo que tuvo éxito” (Kirk y Kutchins, 2002, p. 159). Estos autores critican la metodología empleada para llegar a las conclusiones sobre la fiabilidad del DSM-III y las interpretaciones que se hicieron de los resultados de los análisis estadísticos llevados a cabo en los estudios que se realizaron. La discusión sobre la fiabilidad de los juicios diagnósticos se mantiene de cara a la publicación del DSM-5. Recientemente se han publicado los resultados de los estudios de fiabilidad interjueces como parte del desarrollo del DSM-5. En el caso del trastorno depresivo mayor (utilizando diagnósticos independientes y ciegos), el índice kappa se situaba dentro del rango de fiabilidad “cuestionable” (kappa entre 0.20 y 0.40) (Regier et. al., 2013).

Puesto que la fiabilidad de cada categoría diagnóstica en realidad se pone a prueba mediante la aplicación de instrumentos como las entrevistas diagnósticas (generalmente estructuradas o semiestructuradas) y los cuestionarios diagnósticos, más adelante, en el capítulo 3 de la presente tesis doctoral, se volverá sobre el tema de la fiabilidad de los diagnósticos y se profundizará en el mismo en el contexto del análisis de la fiabilidad de cada uno de esos tipos de instrumentos.

#### **2.1.4. Validez del diagnóstico de los trastornos depresivos**

Las clasificaciones y categorías diagnósticas, además de ser fiables, deben ser válidas. El término “válido” en el contexto de las categorías diagnósticas tiene al menos dos significados. Por un lado, las categorías diagnósticas serían válidas si se demuestra que se corresponden con entidades discretas que tienen límites naturales que les separan de otras entidades (Kendell y Jablensky, 2003), es decir, si las categorías apresan fenómenos reales. Por otro lado, las categorías diagnósticas serían válidas en la medida que las características definitorias de dicha categoría o trastorno proporcionan información útil que no está incluida en la



propia definición de la categoría o trastorno (Spitzer, 2001). Puesto que la mayoría de las categorías diagnósticas del DSM-IV o de la CIE-10 están basadas en la presencia de un síndrome característico, en la medida que se demuestre que la identificación de dicho síndrome permite conocer la etiología, los factores de riesgo, el curso habitual o los correlatos biológicos, psicológicos o sociales del citado síndrome, o ayude a tomar decisiones sobre su prevención o tratamiento, tales categorías diagnósticas serían válidas según esa segunda acepción de validez. Aunque algunos autores consideran imprescindibles distinguir ambas acepciones de validez, reservando este término para la primera y utilizando el término “utilidad” o “utilidad clínica” para la segunda (Kendell y Jablensky, 2003), en el estado actual de la ciencia psicopatológica ambos aspectos están íntimamente ligados, de manera que en gran medida la forma de demostrar que dos categorías diagnósticas reflejan dos entidades distintas es demostrando que tienen diferentes etiologías, factores de riesgo, cursos, correlatos biopsicosociales o respuestas al tratamiento, por lo que, en la práctica, parece muy difícil, sino imposible, distinguir entre validez y utilidad clínica, y, por tanto, en la presente tesis doctoral, se considerarán ambos aspectos dentro del concepto de validez.

Teniendo en cuenta este concepto de validez, es importante señalar que en la actualidad, desde enfoques diferentes, muchos autores critican las deficiencias en la validez de los diagnósticos. Por un lado, se señala que el futuro llevará a hacer diagnósticos sobre bases neuroanatómicas o genéticas que no han podido ser utilizados hasta ahora como el criterio de validación de los diagnósticos (Insel et al., 2010). En esta misma línea, Fink señala que “la heterogeneidad de la población [aquella diagnosticada con trastorno depresivo] es la limitación más seria. Los criterios DSM fallan al intentar distinguir a aquellos con causas psicológicas o sociales de su enfermedad de aquellos con una base biológica (endógena)” (“Symposium. Melancholia: Beyond DSM, Beyond Neurotransmitters”, 2007, p. 136;

el texto entre corchetes ha sido añadido para mayor claridad). Abundando en la misma idea, Shorter (2007, p. 6) afirma que haber creado una entidad denominada depresión mayor, “ha sido una catástrofe nosológica de la que 30 años después no nos hemos recuperado” y aboga por separarla en melancolía y trastornos depresivos no melancólicos.

En un sentido prácticamente contrario, Bentall (2006, p. 242) afirma que: “la pretensión de que la esquizofrenia es una enfermedad específica del cerebro parece muy tenue y parece más bien que ciertos componentes genéticos, junto con leves déficits neurológicos, confieren una vulnerabilidad general a todas las formas de trastornos psiquiátricos graves”, señalando la necesidad de ir más allá de la etiqueta diagnóstica para explicar elementos comunes en los procesos patológicos y poniendo en duda su perdurabilidad como entidades.

Uno de los objetivos de las clasificaciones diagnósticas es proporcionar a los clínicos información sobre el tratamiento más adecuado, y en la medida que lo haga tendrá mayor validez (Mullins-Sweat y Widiger, 2009). Sin embargo, sobre este punto surgen algunas dudas a las clasificaciones actuales. Por ejemplo, en los trastornos depresivos se han encontrado fármacos eficaces para su tratamiento, pero a lo largo de las últimas décadas se ha demostrado que estos mismo fármacos son eficaces para una amplia variedad de trastornos (Andrews et al., 2007). De hecho, el descubrimiento de que, por ejemplo, los fármacos inhibidores de la recaptación de serotonina son igualmente eficaces en trastornos del estado de ánimo, trastornos de ansiedad y trastornos de alimentación, ha puesto en duda la utilidad de los sistemas de clasificación actuales para identificar categorías válidas (Regier et al., 2009). En especial se ha estudiado la semejanza en los efectos del tratamiento con antidepresivos entre el trastorno depresivo mayor y el trastorno de ansiedad generalizada (Furukawa, Watanabe y Omori, 2010).

De forma similar, en la medida en que ayude a anticipar el curso de los problemas del paciente o la duración de los síntomas, el diagnóstico será útil para los profesionales y, por tanto, válido (Mullins-Sweatt y Widiger, 2009) y, precisamente, en el caso de los trastornos depresivos, el **curso de los trastornos** es muy variable. Aunque un alto porcentaje de sujetos tiene tan solo un episodio depresivo mayor de unos meses de duración, se sabe que existe una alta probabilidad de recaída, mayor cuanto más elevado sea el número de episodios previos y cuanto más larga es la duración del mismo (Vázquez y Sanz, 2008). No obstante, aunque el trastorno puede ser crónico y recurrente, en un porcentaje alto de sujetos tampoco sucede de este modo.

La existencia de una **alta comorbilidad de síntomas** pertenecientes a diferentes categorías diagnósticas también ha puesto en duda la propia existencia de dichas categorías de los sistemas de clasificación actuales. La intención original de los autores que desarrollaron los modelos categoriales actuales se puede resumir con la idea de Jaspers: “El principio del diagnóstico médico es que todos los elementos que componen una enfermedad sean caracterizados con un solo diagnóstico... en una sola persona” (citado en Regier et al., 2009, p. 646). Pero esto no es la norma, como han demostrado los estudios epidemiológicos. No solamente existe una coocurrencia entre los trastornos, sino que existe una correlación, puesto que la presencia de más de un trastorno se produce con mayor frecuencia de lo esperable por azar (Eaton, South y Krueger, 2010). Quizá a través de esta alta comorbilidad pueden obtenerse otros modos de organización de los trastornos mentales que expliquen mejor el hecho de que algunos trastornos aparezcan juntos en un sujeto, tanto en un momento dado de su vida como sucesivamente a lo largo de su historia vital. Cabe por ejemplo la posibilidad de que existan factores comunes que expliquen los trastornos. Incluso algunos de los que han defendido los sistemas de clasificación actuales, señalan que la comorbilidad de los trastornos

“puede que esté reflejando una limitación del sistema de clasificación en su capacidad representativa de la complejidad sindrómica” (Pincus et al., 2005, p. 136).

Las críticas sobre la validez del diagnóstico también se han centrado en el elemento clave de la constitución de las categorías diagnósticas actuales, es decir, las semejanzas y diferencias en los síntomas y signos presentes. En concreto, se han puesto en duda las **clasificaciones de síndromes por similitud fenotípica**, como lo son las del DSM y la CIE, que dejan de lado las teorías explicativas, porque dichos sistemas de clasificación puede provocar errores fundamentales (Üstün et al., 2002; Zachar, 2008).

Además, también se ha criticado que en ocasiones **se han utilizado criterios de consenso en lugar de criterios empíricamente validados** para definir las categorías diagnósticas (Eaton, et al., 2010). Algunos cambios en los sistemas de clasificación se producen en base a “la investigación empírica”, pero en otras ocasiones “las categorías propuestas obedecen más a consensos políticos que a razones de peso científico” (Vázquez et al., 2011 p. 99).

Las críticas sobre la validez también se han centrado en el hecho de que, a pesar de que era uno de los objetivos del DSM-III y de sus sucesivas versiones, no se han encontrado hasta ahora **marcadores biológicos** para el diagnóstico de los distintos trastornos mentales. Por ejemplo, se han encontrado pistas de marcadores genéticos en el cromosoma 18 para la depresión, pero a la vez, es el mismo cromosoma en el que se han encontrado marcadores para la predisposición a la esquizofrenia (Jablensky y Kendell, 2002). Existen también intentos de validar los instrumentos de evaluación diagnóstica a través de los estudios de **herencia genética**. En este sentido, se ha encontrado una mayor probabilidad de padecer esquizofrenia o depresión mayor entre los familiares de primer grado de las

personas con estos trastornos. Sin embargo, estos estudios, a su vez, pueden dar más pistas acerca del continuo que existe entre trastornos y no sobre la validez de las categorías diagnósticas específicas, puesto que en realidad lo que parece que se hereda sería la predisposición a los trastornos del “espectro depresivo” o del “espectro psicótico”, y no la predisposición concreta a la depresión mayor o a la esquizofrenia. Más aún, algunos eventos ambientales contribuyen de igual medida al origen de varios síndromes diferentes (Maj, Gaebel, López-Ibor y Sartorius, 2002). En esta misma línea, los estudios sobre la base genética de la ansiedad generalizada y de la depresión mayor, indican que estos trastornos “comparten algunos, si no la mayoría, de sus factores de riesgo genéticos” y este hecho también se encuentra en relación con otros trastornos de ansiedad como el de pánico (Hettema, 2010, p.21).

Probablemente, las dificultades para obtener marcadores biológicos para la mayoría de los trastornos mentales, incluyendo los trastornos depresivos, tienen que ver en parte con la heterogeneidad de los síndromes y pacientes a los que se les asigna una misma categoría diagnóstica, heterogeneidad que ha sido criticada desde el punto de vista de la validez de dichas categorías. Así, con el objetivo de conocer el curso de los trastornos, las respuestas al tratamiento y otras facetas de los trastornos mentales, **se ha considerado el diagnóstico como un elemento homogéneo**, pero las facetas que componen los trastornos actuales no cumplen esa característica, lo que puede enturbiar los resultados de los análisis estadísticos que tratan de validar dichos trastornos (Meehl, 2010), de manera que el uso del trastorno depresivo como una variable categorial en investigación puede ser problemática.

### **Modelos categoriales frente a dimensionales para los trastornos depresivos**

Los datos epidemiológicos disponibles en estos momentos encajan mejor con las aproximaciones dimensionales que con los modelos categoriales. La sintomatología de los trastornos depresivos, según Andrews et al. (2007, p. 43):

“se distribuye de manera dimensional con mayor o menor intensidad, persistencia en el tiempo, cambios en los estados habituales, niveles de malestar y deterioro funcional. Es tarea de los clínicos juzgar si la gravedad, la duración, la anormalidad, el malestar o el deterioro de los síntomas del sujeto justifican la asignación de una categoría diagnóstica, es decir, los síntomas exceden un hipotético umbral en este espacio multidimensional, mediante el cual se puede justificar un diagnóstico y conseguir una intervención beneficiosa. Todo esto parece complejo, y lo es.”

En el caso del trastorno depresivo mayor, aunque existe una relación lineal entre el número de síntomas y el grado de deterioro, no existe evidencia de ninguna discontinuidad natural que pueda establecer una frontera entre la normalidad y el trastorno, estimado actualmente en el uso del punto de corte de 5 síntomas sobre 9 como un umbral diagnóstico (Andrews et al., 2007). En esta misma línea, según la guía NICE (2010, p. 23):

“Dada la dificultad práctica y la incertidumbre inherente a la hora de decidir los umbrales de gravedad y disfuncionalidad significativos de los síntomas, no existe una discontinuidad natural entre los síntomas depresivos subclínicos y la depresión ‘mayor leve’ en la práctica clínica habitual”.

Los pacientes con sintomatología depresiva tienen un grado alto de deterioro en el funcionamiento psicológico y social, aún sin cumplir los criterios para el diagnóstico de depresión mayor (Kessler, Merikangas, Berglund y Eaton, 2003). Además, los pacientes con sintomatología depresiva tienen una mayor probabilidad de tener episodios posteriores de depresión más graves (Cuijpers y Smit, 2004 y Cuijpers, Smit y Willemse, 2005).

El propio DSM-IV recoge en las especificaciones del trastorno depresivo mayor distintas categorías de gravedad en función del número de síntomas y la disfuncionalidad o deterioro que provocan, como reconocimiento a la dimensionalidad de este trastorno, aportando además la categoría en estudio de trastorno depresivo menor para aquellos pacientes con sintomatología subumbral. Como ya se ha comentado en el primer capítulo de esta tesis, la CIE-10 también clasifica la gravedad del trastorno depresión, aunque sólo en función del número de síntomas del paciente y no por el grado de deterioro que provocan.

Las propuestas empíricamente validadas a través de mediciones dimensionales de la sintomatología (p. ej., Krueger, Watson y Barlow, 2005), tratan de superar las dificultades de los modelos categoriales. En el caso de los trastornos depresivos, las conclusiones coinciden con lo comentado hasta el momento sobre la continuidad entre la normalidad y el trastorno (Slade, 2007; Slade y Watson, 2005, en Vázquez y Sanz, 2008). Además, se considera que algunos trastornos de ansiedad (claramente el trastorno de ansiedad generalizada y, en menor medida, el trastorno por estrés postraumático) encajan mejor dentro del espectro de los problemas de los trastornos del estado de ánimo. Estos resultados son muy homogéneos transculturalmente. Estudios a gran escala efectuados en EE. UU., Australia y en una docena más de países, demuestran una gran similitud en este patrón diagnóstico (Krueger y Markon, 2006).

Por otro lado, se ha encontrado que pequeños cambios en el umbral diagnóstico producen grandes cambios en el porcentaje de diagnósticos de los trastornos depresivos (Karlsson, Marttunen, Karlsson, Kaprio y Hillevi, 2010). Determinar ese punto de corte será positivo para poder obtener una buena fiabilidad en el diagnóstico. Mullins-Sweatt y Widiger (2009) consideran que los umbrales diagnósticos actuales no están en su punto óptimo para muchas situaciones sociales o clínicas. Por tanto, cumplir o no los criterios diagnósticos de los trastornos depresivos sería en estos momentos poco relevante para la toma de decisiones en la práctica clínica y en buena medida para la investigación de los procesos subyacentes a los trastornos depresivos.

Para algunos autores, las formas depresivas que se sitúan por debajo del umbral diagnóstico “son particularmente importantes por su prevalencia, significación clínica y coste sociosanitario”; pero una vez analizada la bibliografía, señalan la necesidad de investigación al respecto, porque se desconoce su pronóstico, los tratamientos indicados y su eficacia (Pincus et al., 2005). Cuando no se cumplen los criterios diagnósticos suficientes, pero sí se considera que es necesaria una intervención terapéutica, por ejemplo en una persona con sintomatología depresiva subumbral, las decisiones diagnósticas se consideran confusas y se analizan poco (Üstün et al., 2002). Quizá no se han estudiado precisamente porque la investigación se ha centrado durante las últimas décadas en aquellos aspectos que incluyen solamente las etiquetas diagnósticas con el trastorno depresivo mayor. Por esta razón, en los futuros sistemas diagnósticos se ha recomendado facilitar la evaluación de la sintomatología leve, así como la comorbilidad entre depresión, ansiedad y trastornos somatomorfos, con una aproximación dimensional (Hanel et al., 2009).



### **El concepto de significación clínica**

Siguiendo el manual DSM-IV-TR, no basta con cumplir una serie de síntomas sino que éstos deben tener un grado de “significación clínica”: “En este manual cada trastorno mental es conceptualizado como un síndrome o un patrón comportamental o psicológico de significación clínica, que aparece asociado a un malestar (p. ej., dolor), a una discapacidad (p. ej., deterioro en una o más áreas de funcionamiento) o a un riesgo significativamente aumentado de morir o de sufrir dolor, discapacidad o pérdida de libertad” (APA, 2002, p. XXI). Wakefield y First (2005) señalan que el concepto de significación clínica encierra un argumento circular “porque la decisión sobre si un síndrome o patrón resulta clínicamente significativo depende de si es o no considerado un trastorno” (p. 45). En cualquier caso, el grado de deterioro y malestar está asociado a la gravedad de la sintomatología depresiva y los síntomas de la depresión mayor casi siempre suponen deterioro y malestar (Wakefield, Schmitz y Baer, 2010). De manera que, al igual que se decía respecto a la gravedad de la sintomatología, un criterio algo más exigente para determinar si una serie de síntomas interfieren en la vida cotidiana puede llegar a reducir a la mitad las estimaciones del número de personas con depresión o ansiedad en la población general (Narrow, Rae, Robins y Regier, 2002).

### **Consideraciones finales sobre la validez y utilidad del diagnóstico de los trastornos del estado de ánimo**

Se han comentado anteriormente algunas dificultades para la conceptualización de los trastornos mentales, especialmente de los trastornos depresivos. La aplicación de análisis taxonómicos a la psicopatología está relativamente en sus etapas tempranas; en particular esta metodología ha sido utilizada en el estudio de la depresión mayor en muy pocos estudios y con

resultados contradictorios y no concluyentes (Joiner et al., 2005). Es posible que con esta metodología se encuentren modelos psicopatológicos que proporcionen una visión más ajustada a los datos actuales.

La influencia de las categorías diagnósticas en la práctica clínica, en la investigación, en la definición de las políticas sanitarias y en la concepción social de la salud mental es enorme. Las clasificaciones configuran el modo en el que se ve el mundo y a qué se dirige la atención, pero en estos momentos existen dudas acerca de la capacidad de estos sistemas para explicar la información disponible. Para los instrumentos de evaluación de los trastornos depresivos será una tarea complicada detectar los sujetos que se agrupan bajo estas etiquetas siendo grupos tan heterogéneos, situados dentro de una variable continua y con una gran comorbilidad sintomatológica. Sin embargo, en este momento no se dispone de una alternativa bien definida que pueda sustituir a corto plazo a las actuales clasificaciones diagnósticas.

La falta de información puede hacer que los clínicos y la sociedad en su conjunto asuman las etiquetas diagnósticas como elementos explicativos de la sintomatología de los pacientes. Eso nos llevaría al argumento circular: *está usted deprimido porque tiene depresión*. Al contrario, en la búsqueda de una mayor eficacia en los tratamientos, el diagnóstico debe servir para mejorar nuestros modelos explicativos y ajustarlos lo máximo posible a cada población o incluso a cada individuo. “Una vez explicado correctamente por qué aparecen en una persona alucinaciones, delirios, trastornos del pensamiento y demás, no quedará por explicar algo como la esquizofrenia” (Bentall, 2006, p. 42 y 43).

La discusión sobre la naturaleza del diagnóstico de los trastornos depresivos es fundamental para entender su utilidad y las dificultades que conlleva el estudio de la fiabilidad y validez de los instrumentos diagnósticos. Las consecuencias de

estar utilizando unas categorías que están seriamente en duda son muy importantes y la utilidad del diagnóstico mismo como variable predictiva está puesta en duda. Sin embargo, “los endeble fundamentos de los sistemas diagnósticos no los hacen irrelevantes” y “son cada vez más utilizadas, y su uso tiene enormes repercusiones económicas y sociales” (Vázquez et al., 2011, p. 93).

De hecho, durante la última década se ha promovido la utilización de métodos de detección de los trastornos depresivos con el objetivo de generalizar el tratamiento en la población. Al respecto, Horwitz y Wakefield (2007) advierten que la implantación de la medicación y de los métodos de *screening* masivos es incompatible con un análisis contextual del sujeto y las circunstancias en las que esos síntomas se producen, patologizando la tristeza hasta convertirlo en algo que es inaceptable para la sociedad. La guía NICE (2010) recomienda no prescribir fármacos para los sujetos con depresiones leves debido a la baja relación coste/eficacia, pero este tipo de tratamientos se continúan implantando masivamente en la actualidad (Pinto-Meza, Fernández, Serrano-Blanco y Haro, 2008, en Cano Vindel, Martín Salguero, Mae Wood, Dongil y Latorre, 2012).

## **2.2. El diagnóstico de los trastornos depresivos**

Los criterios de la depresión mayor permanecen bastante estables desde hace más de 30 años, lo que ha provocado que los conocimientos de varias generaciones de psicopatólogos, psiquiatras y psicólogos clínicos tengan un mayor conocimiento sobre la sintomatología depresiva. Anteriormente se han comentado algunos problemas con el diagnóstico de los trastornos mentales, en especial de los trastornos depresivos: la dificultad para establecer un umbral diagnóstico, la heterogeneidad de la sintomatología en los sujetos, la alta comorbilidad o la ausencia de marcadores biológicos. A estos problemas se podría añadir otro más relacionado con la definición de los signos y síntomas que forman parte de los

criterios diagnósticos de los trastornos depresivos, ya que dichas definiciones podrían ser más precisas. Por ejemplo, en la CIE-10 y, en menor medida, en el DSM-IV no se define del modo correcto la tristeza o la anhedonia (Vázquez, et al., 2011). Parece, por tanto, necesario tratar de explicitar a continuación cada uno de estos síntomas con más detalle, así como el diagnóstico diferencial, antes de detenerse en describir los principales instrumentos disponibles para su evaluación.

### 2.2.1. El diagnóstico del trastorno depresivo mayor

Dentro de los trastornos del estado de ánimo se distinguen cuatro tipos de episodios: episodio depresivo mayor, episodio maníaco, episodio hipomaniaco y episodio mixto. Para el diagnóstico de trastorno depresivo mayor al menos se ha debido presentar un **episodio depresivo mayor**, sin la presencia de episodios maníacos, mixtos o hipomaniacos. Los criterios para dicho episodio, según el DSM-IV, son los siguientes:

**A)** Para cumplir este primer criterio deben estar presentes 5 o más síntomas de un listado de 9 síntomas posibles, durante un periodo de al menos dos semanas consecutivas, y que tales síntomas representen un cambio con respecto a la actividad previa (para la valoración de esta sintomatología se recomienda que el evaluador centre su atención específicamente en las dos semanas en las que el paciente se ha sentido peor durante el último mes). Al menos uno de los síntomas debe ser el criterio A.1 o el criterio A.2, que deben cumplirse casi cada día, la mayor parte del día.

A.1) El **estado de ánimo depresivo** debe ser evaluado mediante descripciones por parte del paciente, por sus familiares o inferido por su expresión no verbal durante las entrevistas. Los pacientes depresivos utilizan expresiones como “deprimido, triste, desesperanzado, desanimado o «como en un pozo»” (APA, 2002, p. 327). En cambio, algunos pacientes manifiestan incapacidad para sentir

emociones o señalan simplemente que están ansiosos. Las ganas de llorar, el propio llanto, la mirada baja y otras expresiones de tristeza pueden ser considerados como indicadores de bajo estado de ánimo. Además, el hecho de que el paciente señale que él no se siente triste no debe considerarse necesariamente como prueba inequívoca de la inexistencia de este síntoma; este criterio debe valorarse como presente si existen observadores de su entorno que señalen la presencia de los mencionados signos externos de bajo estado de ánimo (First et al., 1999). Además, algunas personas tienen sensación de vacío o incluso dificultad para sentir emociones, “de ahí que la *incapacidad para llorar* sea un síntoma que indica una mayor gravedad de la depresión” (Vázquez y Sanz, 2008, p. 241).

Aunque más frecuente en los niños y adolescentes, los adultos pueden sentirse irritables en lugar de tristes. De hecho, aproximadamente el 50% de los sujetos con depresión mayor siente irritabilidad (Fava et al., 2010). En este sentido el DSM-IV indica que pueden aparecer “ira persistente, tendencia a responder a los acontecimientos con arranques de ira o insultando a los demás, o sentimiento exagerado de frustración por cosas sin importancia” (APA, 2002, p. 327).

Lewinsohn, Petit, Joiner y Seeley (2003) señalan que el criterio del estado de ánimo deprimido está presente en el 95% de los pacientes depresivos, es la queja más frecuente que presentan los sujetos y es un criterio relativamente fácil de detectar.

Por otro lado, la tristeza no siempre debe ser considerada un síntoma de trastorno mental. De hecho todas las emociones juegan un papel fundamental en nuestra capacidad para adaptarnos a nuestro medio. Aunque durante las últimas décadas se ha comenzado a estudiar el papel de las emociones positivas, con frecuencia se olvida que la tristeza también tiene un papel útil desde el punto de vista evolutivo. Bonnano, Goorin y Coifman (2008) señalan que esta emoción nos

orienta el procesamiento de la información de manera detallada, genera valoraciones más precisas, mejora la toma de decisiones, mejora la disposición a resolver problemas que consuman más tiempo y nos permiten comunicarnos de manera eficaz con los demás. Sin embargo la tristeza prolongada o con una intensidad mayor de lo esperado para una situación determinada puede ser un síntoma de un trastorno depresivo.

**A.2) La pérdida de interés o de la capacidad para experimentar placer (anhedonia).** Se ha descrito como un “bloqueo del refuerzo de gratificación de los estímulos generalmente reforzadores” (Rado, 1956, en Berrios, 2008). Los sujetos que cumplen este criterio suelen haber perdido la capacidad para disfrutar de las actividades que antes realizaban de manera cotidiana, de sus aficiones y en general de las actividades que solían resultarle placenteras. Este criterio está unido a la apatía y la pérdida de interés. La entrevista CIDI 2.1 señala que para el cumplimiento de este criterio, la pérdida de interés debe aparecer “en la mayoría de las cosas, como el trabajo, los pasatiempos y otras cosas que normalmente disfrutaba” (OMS, 1997). Con frecuencia estas dificultades vienen acompañadas del abandono de las relaciones sociales así como una disminución en el deseo sexual.

Según se describe en el manual DSM-IV, para cumplir el criterio A, deben aparecer además al menos cuatro (tres en caso de que se cumplieran los dos anteriores) de los siguientes síntomas (desde A.3 a A.9):

**A.3) Pérdida o ganancia del peso o cambios en el apetito.** La forma más común es la pérdida del peso, por lo que debe tomarse en consideración si el paciente está realizando una dieta. Vázquez y Sanz (2008) señalan que existe una cierta polémica sobre el criterio de aumento de peso o apetito, desde su inclusión como criterio en el DSM-III debido a su dificultad para ser valorado. Por ejemplo, a veces el paciente puede que tenga poco apetito, pero coma dulces y no pierda

peso, que tenga una enfermedad o una medicación que le haga variar de peso o apetito, que coma como siempre pero no disfrute de la comida o que tenga altibajos en el apetito. Estos autores añaden que hay dos preguntas muy útiles para la valoración de este criterio, igualmente válidas para otros síntomas como los cambios en el sueño o los problemas psicomotores: “¿Es esta condición diferente a lo que sucede normalmente en esta persona?” y “¿El paciente lo vive como un problema?”.

El DSM-IV, en el caso de la valoración del cambio de peso, propone una variación de un 5% en un mes como un elemento significativo (APA, 2002). En la entrevista CIDI 2.1 se señala como definición del síntoma que la variación debe ser de 1 kilo por semana arriba o abajo durante varias semanas o más/menos apetito de lo usual durante todos los días (OMS, 1997).

En la entrevista SCID-I se llama la atención sobre la posibilidad de que la pérdida de peso no esté acompañada por una disminución del apetito o de la ingesta, porque podría ser una señal de una enfermedad física (First et al., 1999).

**A.4) Cambios en las pautas de sueño**, ya sea insomnio o hipersomnia. Suele caracterizarse por el insomnio medio (despertarse a media noche y tener dificultades para volver a conciliar el sueño durante más de media hora), el insomnio inicial (tardar más de media hora en quedarse dormido) y el despertar temprano (entre 1 y 3 horas antes de la hora habitual) (Vázquez y Sanz, 2008). Las entrevistas CIDI 2.1 (OMS, 1997) y la SCID-I (First et al., 1999) señalan como criterio de significación clínica de este síntoma que el paciente tarde 2 horas más en conciliar el sueño o se despierte 2 horas antes de lo habitual.

Menos común es el aumento de horas de sueño, tanto nocturnas como diurnas, para el que no hay un criterio definido, tanto para las horas que se duerman durante la noche como para valorar la existencia de un estado de mayor

somnolencia de lo habitual. Vázquez y Sanz (2008) señalan que debe analizarse además el estado de somnolencia del paciente durante las 24 horas, así como si el sueño es reparador y el tiempo que pasa en la cama. Pueden aparecer dificultades para valorar este diagnóstico si la persona está tomando medicación, tiene turnos rotatorios o presenta dolores.

**A.5) Agitación o enlentecimiento psicomotores.** La agitación se define como el aumento en los movimientos, por ejemplo, paseos, pellizcos o frotarse las manos. Según la CIDI 2.1 (OMS, 1997) se define como “dificultades para mantenerse quieto o mostrar un movimiento continuo”. Estos aspectos deben ser detectados por el evaluador o las personas cercanas al paciente, y no debe ser tan solo una percepción subjetiva del propio sujeto. Por ejemplo, en la entrevista SCID-I (First et al., 1999) se pregunta si este síntoma “¿Fue tan acusado como para que los demás se dieran cuenta?”. El enlentecimiento se caracteriza por un lenguaje expresivo más lento, aumento en la latencia de respuesta, bajo volumen de voz, menos inflexiones o mutismo.

**A.6) Es habitual encontrar en las personas depresivas fatiga o pérdida de energía.** Se produce agotamiento en las tareas que requieren habitualmente poco esfuerzo, así como un mayor gasto de tiempo y una menor eficacia en su ejecución. Se debe distinguir entre pérdida de energía, pérdida de interés y las dificultades derivadas de los problemas de sueño (Vázquez y Sanz, 2008).

**A.7) Sentimiento de inutilidad o culpa excesiva.** El sentimiento de culpa puede aparecer como una manifestación de un sentimiento exagerado de responsabilidad ante situaciones adversas o pequeños errores del pasado. Por su parte, el sentimiento de inutilidad excesiva hace referencia a la percepción negativa de uno mismo. Vázquez y Sanz (2008, p. 251) señalan que no basta con “la presencia de reproches por estar enfermo, sentirse incapaz o inválido, sino que se



trata de algo más intenso, reiterado y que supone una preocupación casi constante para el paciente”. En la entrevista SCID-I (First et al., 1999) se incluyen las preguntas “¿Qué pensaba de sí mismo?” “¿Pensaba que era un inútil?”, con la advertencia de que debe ser codificado como negativo “si se trata tan solo de escasa opinión de sí mismo” (First et al., 1999, p. XX). Por su parte, la entrevista CIDI 2.1 (OMS, 1997) propone las siguientes preguntas: “¿Sentía que no era tan bueno(a) o era inferior a otras personas?” y “¿Tenía tan poca confianza en sí mismo(a) que no se atrevía a opinar?”.

**A.8) Disminución de la capacidad para pensar, tomar decisiones o concentrarse.** Es frecuente entre las personas con otros síntomas depresivos las quejas sobre dificultades para concentrarse en tareas y dificultades de memoria. Vázquez y Sanz (2008) sugieren para su evaluación preguntas del tipo “¿Es capaz el paciente de concentrarse viendo la televisión, seguir una conversación o concentrarse en el trabajo?”, dejando claro que se trata de evaluar la concentración más que el interés. La SCID-I por su parte propone las siguientes preguntas: “¿Tenía dificultades para pensar o concentrarse?” y “¿Con qué interferían?” (First et al., 1999). Como en los restantes síntomas, se debe tratar de valorar si existe un cambio con respecto al rendimiento anterior del paciente.

Por su parte, la entrevista CIDI 2.1 (OMS, 1997) hace hincapié en el deterioro de la capacidad para concentrarse, de manera que una persona con altas capacidades puede leer, escuchar la radio o mirar la televisión o películas, pero no hacer cosas que antes hacía en su trabajo o en sus aficiones.

En los criterios A.3) hasta el A.8) es fundamental tener en cuenta que deben estar presentes casi cada día, durante una gran parte del día, durante un periodo mínimo de dos semanas consecutivas.

**A.9)** El **pensamiento recurrente de muerte o ideación suicida** varía entre pensamientos relacionados con la posibilidad de descansar definitivamente o la idea de que los demás estarían mejor sin ellos, hasta la planificación detallada o el intento suicida. Aunque se recomienda la evaluación de elementos como la duración, la frecuencia y el grado de estructuración de estos pensamientos suicidas, no se han encontrado hasta ahora predictores eficaces para detectar cuándo se producirá un intento suicida (APA, 2002). No obstante, es un síntoma que debe valorarse de manera sistemática en los pacientes, en especial cuando tienen sintomatología depresiva. Para ello, es positivo hacerlo con sensibilidad y empatía, pero sin miedo a que el hecho de preguntar acerque a la persona a quitarse la vida, ya que éste es un mito que no se corresponde con la realidad (Joiner, 2010). Se han propuesto preguntas como “¿Estaba tan mal las cosas que pensaba mucho en la muerte o que sería mejor estar muerto? ¿Pensaba en hacerse daño a sí mismo?” (First et al., 1999).

**B)** Los síntomas no deben cumplir criterios para un episodio mixto, para lo que deberemos valorar hasta que punto el paciente ha compartido al menos durante una semana criterios depresivos y maníacos o hipomaniacos.

**C)** Los síntomas provocan malestar clínicamente significativo o deterioro social, laboral o de otras áreas importantes de la actividad del individuo. En el DSM-IV se especifica que, por ejemplo, el paciente puede perder su capacidad para relacionarse o trabajar. La entrevista SCID-I propone la pregunta: “¿La (depresión/o las propias palabras que utiliza el sujeto) le hizo más difícil trabajar, atender las tareas de la casa o relacionarse con otras personas?” (First et al., 1999).

**D)** La sintomatología no debe ser el resultado de una enfermedad médica o debido al consumo de sustancias. El DSM-IV destaca que debe estar asociado claramente la enfermedad física o el consumo de sustancias con la aparición de los

síntomas que definen el episodio depresivo, puesto que la depresión puede tener exactamente las mismas características independientemente de su origen. Vázquez y Sanz (2008) señalan que es el análisis inicial más importante ante un diagnóstico de trastorno del estado de ánimo. Quizá un 10 o un 20% de las depresiones o distimias están causadas por factores orgánicos de diversos tipos (Rush, 1990, en Vázquez y Sanz, 2008), por lo que estarían incorrectamente diagnosticadas y, según el DSM-IV, se deberían diagnosticar como trastorno del estado de ánimo debido a enfermedad médica (bien el subtipo con episodio similar al depresivo mayor o el subtipo con síntomas depresivos) o como trastorno del estado de ánimo inducido por sustancias (subtipo con síntomas depresivos). Por supuesto, para realizar estos últimos diagnósticos es importante, tal y como señala el DSM-IV, que en la historia clínica, la exploración física o los hallazgos de laboratorio existan pruebas de que la alteración es la consecuencia directa de una enfermedad médica o del consumo de sustancias (drogas o medicamentos). Esta relación se puede concretar cuando hay una asociación temporal entre la aparición, exacerbación o disminución de los síntomas depresivos y la aparición, exacerbación o disminución de la enfermedad a la que está asociada o la ingesta o abstinencia de una sustancia.

**E)** Además, los síntomas no se explican mejor por la presencia de un duelo, o si existe duelo, los síntomas persisten durante más de dos meses o se caracterizan por una acusada incapacidad funcional, preocupaciones mórbidas de inutilidad, ideación suicida, síntomas psicóticos o enlentecimiento psicomotor. Como se ha comentado anteriormente, los criterios del DSM-IV-TR para el diagnóstico de depresión mayor incluyen que si la sintomatología depresiva aparece como reacción ante la muerte de un ser querido, el diagnóstico será positivo en el caso de que dicha sintomatología dure más de 2 meses.

Para valorar bien la presencia o ausencia de un episodio depresivo mayor, la evaluación de los criterios diagnósticos anteriormente mencionados debe hacerse teniendo en cuenta diversos factores que pueden afectar a la fiabilidad y validez de la información diagnóstica. Por ejemplo, el DSM-IV-TR destaca que en la cultura latina y mediterránea la presentación de síntomas depresivos está asociada a los “nervios y las cefaleas” y, en general, señala la necesidad de ajustar el proceso diagnóstico a la cultura del sujeto (APA, 2002). Por otro lado, los propios síntomas de los pacientes pueden afectar a la correcta valoración de la información diagnóstica. Como antes se ha comentado, los pacientes depresivos sienten apatía, falta de interés y baja motivación, por lo que durante la evaluación psicológica pueden aparecer sesgos cognitivos y emocionales que tienden a la negatividad. El DSM-IV también señala que la información puede verse comprometida por la falta de concentración, el deterioro en la memoria, la negación y la justificación de los síntomas.

La presencia de un episodio depresivo mayor no tiene asignado ningún código en el DSM-IV, puesto que puede formar parte de un trastorno ciclotímico o de un trastorno depresivo mayor o estar asociado a otro trastorno. Según el DSM-IV, el trastorno depresivo mayor se caracteriza por la presencia de al menos un episodio depresivo mayor, que no debe haberse producido ni por una enfermedad médica ni debido al consumo de sustancias y que no se solapa con la aparición de esquizofrenia, de un trastorno esquizofreniforme, de un trastorno delirante o de un trastorno psicótico no especificado. El trastorno esquizoafectivo se asigna en el caso de que el sujeto tenga alucinaciones o delirios durante un periodo de al menos dos semanas. En el caso de la esquizofrenia, el trastorno delirante o el trastorno psicótico no especificado, la distinción suele ser clara, aunque con frecuencia aparecen síntomas depresivos en los trastornos de este espectro. Especialmente la

esquizofrenia de tipo catatónico tiene una manifestación intensa de síntomas negativos.

Para ser diagnosticado con un trastorno depresivo mayor, el sujeto no debe haber cumplido los criterios para un episodio maniaco o hipomaniaco. Este punto es fundamental y complicado de valorar, puesto que requiere una valoración específica y retrospectiva por parte del evaluador, e implica que el sujeto informe de posibles situaciones pasadas que con frecuencia no son egodistónicas. Los episodios depresivos del trastorno depresivo mayor son indiferenciables de los transcurridos en personas que cumplen los criterios de episodio de manía o hipomanía, de manera que entre el 10% y el 15% de los sujetos diagnosticados con depresión unipolar acabarán siendo diagnosticados de bipolares (Olie y Cols, 1992, en APA, 2002).

Por último, al realizar el diagnóstico de trastorno depresivo mayor según el DSM-IV se pueden tener en cuenta las siguientes **especificaciones** (véase la Tabla 2.1.) con el objetivo de describir el episodio depresivo mayor actual (o para el episodio depresivo mayor más reciente si en ese momento no se cumplen los criterios para un episodio depresivo mayor):

**Tabla 2.1.** Especificaciones para el trastorno depresivo mayor según el DSM-IV-TR (APA, 2002)

<b>Codificar el estado del episodio actual o más reciente</b>	<b>Especificar (para el episodio actual o el más reciente):</b>
.0 Leve	Crónico
.1 Moderado	Con síntomas catatónicos
.2 Grave sin síntomas psicóticos	Con síntomas melancólicos
.3 Grave con síntomas psicóticos	Con síntomas atípicos
.4 En remisión parcial/en remisión total	De inicio en el postparto
.9 No especificado	

El manual clasifica la **gravedad del episodio depresivo actual** en *episodio depresivo leve, moderado, grave sin síntomas psicóticos o grave con síntomas psicóticos*. Existe un amplio apoyo sobre la estabilidad temporal de la clasificación de la gravedad del trastorno depresivo, la cual es más estable cuanto más grave es el diagnóstico del sujeto (Kessing, 2007), y esto a pesar de que esta escala de valoración de la gravedad es arbitraria. De hecho, el DSM-IV describe estas categorías de una manera un tanto difusa (APA, 2002, p. 385):

“Los episodios *leves* se caracterizan por la presencia de sólo cinco o seis síntomas depresivos y por una incapacidad leve o por una capacidad de función normal, pero a costa de un esfuerzo importante e inhabitual. Los episodios *graves sin síntomas psicóticos* se caracterizan por la presencia de la mayoría de los síntomas de los criterios y por una incapacidad evidente y observable. Los episodios *moderados* tienen una gravedad intermedia entre los leves y los graves”.

Según la metodología utilizada para la detección de la gravedad de los síntomas los resultados son diferentes (Kessler, Berglund, Demler, y Jin 2003; Demyttenaere, 2004), también cuando se utilizan cuestionarios como el BDI-II o el PHQ-9. Teniendo en cuenta la importancia que durante los últimos años se está otorgando a la dimensionalidad del trastorno depresivo, recogida mediante estas especificaciones sobre su gravedad, sería interesante la realización de estudios respecto a la fiabilidad y validez de las mismas.

El **trastorno depresivo mayor crónico** se diagnostica cuando los criterios se han cumplido durante los dos últimos años, mientras que la especificación del **trastorno depresivo mayor con síntomas catatónicos** “está caracterizado por una marcada alteración psicomotora que puede consistir en inmovilidad motora,

actividad motora excesiva, negativismo extremo, mutismo, peculiaridades del movimiento voluntario, ecolalia o ecopraxia” (APA, 2002, p. 466).

Se utiliza la especificación con **síntomas melancólicos** cuando aparece una marcada anhedonia, reflejada en una pérdida en todas o casi todas las actividades o en una falta de reactividad emocional ante estímulos habitualmente placenteros (el estado anímico de la persona no mejora ni siquiera cuando sucede algo bueno), junto con tres o más de un conjunto de seis síntomas, la mayoría de ellos de tipo vegetativo: inhibición o agitación psicomotora, anorexia significativa, despertar precoz, peor estado de ánimo por la mañana, culpabilidad excesiva o una cualidad diferente del estado de ánimo depresivo (APA, 2002). Una vez más los autores del DSM-IV no introducen consideraciones etiológicas sobre este subtipo de trastorno depresivo, sino más bien una caracterización descriptiva.

El **trastorno del estado de ánimo depresivo con síntomas atípicos** se refiere a dos conceptos diferenciados. Por un lado al hecho de que el paciente no responde con un aumento del estado de ánimo ante variaciones contextuales positivas y por el otro al cumplimiento de síntomas vegetativos “inversos” (hipersomnia, aumento del apetito o del peso), hipersensibilidad al rechazo social o sentimientos de abatimiento físico (APA, 2002).

En una revisión sobre la evaluación de la depresión basada en la evidencia, se señalan los datos a favor y en contra de la existencia taxonómica de la depresión melancólica y de la depresión atípica, destacándose que la información es más uniforme a favor de la validez de la depresión atípica que a favor de la depresión malancólica (Joiner et al., 2005).

Por su parte, el **trastorno depresivo de inicio en el posparto** es una especificación que debe asignarse si el inicio del episodio depresivo es dentro de las cuatro primeras semanas después del parto.

Estas especificaciones no tienen en el DSM-IV una codificación específica. La discusión continuará durante los próximos años de cara a la publicación del DSM-5 en cuya página oficial (APA, 2011) no hay hasta la fecha notificación alguna sobre, por ejemplo, las especificaciones diagnósticas de depresión atípica y depresión melancólica.

### **2.2.2. El diagnóstico del trastorno distímico**

La introducción de este trastorno en el DSM-III pretendía recoger aquellos síntomas depresivos de larga duración, pero con sintomatología más leve. Mientras el episodio depresivo mayor se caracteriza por cumplir una serie de síntomas “la mayor parte del día, casi cada día”, durante un periodo de al menos dos semanas, los criterios del trastorno distímico deben estar presentes “la mayoría de los días en los últimos dos años”. Los criterios recogidos en el DSM-IV para el trastorno distímico son ocho (APA, 2002):

**A)** Estado de ánimo crónicamente depresivo la mayor parte del día, la mayoría de los días, manifestado por el sujeto u observado por los demás, durante al menos dos años. El DSM-IV señala que “los sujetos pueden señalar pérdida de intereses y aumento de la autocrítica, viéndose a menudo a sí mismos como poco interesantes o inútiles” (APA, 2002, p. 422). Señala además la sensación del sujeto de haber incorporado los síntomas a la vida cotidiana, por lo que puede no referir los síntomas si no es preguntado directamente por ellos.

**B)** Presencia, mientras se está deprimido, de dos (o más) de los siguientes síntomas:

- (1) Pérdida o aumento de apetito. Se debe tener en cuenta que a diferencia del trastorno depresivo mayor, en este caso no se hace referencia a la pérdida o aumento de peso que debe alcanzarse.



(2) Insomnio o hipersomnia.

(3) Falta de energía o fatiga.

(4) Baja autoestima. A diferencia de lo que sucede en el trastorno depresivo mayor, para cumplir este criterio es suficiente con que el sujeto se infravalore o se sienta fracasado.

(5) Dificultades para concentrarse en tareas o para tomar decisiones.

(6) Sentimientos de desesperanza.

**C)** Durante el período de dos años (un año en niños y adolescentes) de la alteración, el sujeto no ha estado libre de síntomas de los Criterios A y B durante más de 2 meses seguidos.

**D)** No ha habido ningún episodio depresivo mayor durante los primeros 2 años de la alteración. En la entrevista SCID-I se pregunta al sujeto “¿Empezó gradualmente o con una época de depresión?” como método para valorar este criterio (First et al., 1999). Una vez pasado este periodo temporal, si el paciente cumple los criterios de depresión mayor, se le diagnosticará ambos trastornos (depresión doble).

**E)** Al igual que en el trastorno depresivo mayor, nunca ha habido un episodio maniaco o hipomaniaco, ni se han cumplido criterios para un trastorno ciclotímico.

**F)** La alteración no aparece exclusivamente en el transcurso de un trastorno psicótico crónico, como son la esquizofrenia o el trastorno delirante.

**G)** Los síntomas no son debido a los efectos fisiológicos directos de una sustancia o una enfermedad médica, en este caso, las mismas enfermedades que se señalan para el trastorno depresivo mayor.

**H)** Los síntomas causan un malestar clínicamente significativo o deterioro social, laboral o de otras áreas importantes de la actividad del individuo.

Por último, al realizar el diagnóstico de trastorno distímico según el DSM-IV, se debe especificar si el inicio del trastorno ha sido temprano (antes de los 21 años) o tardío (más allá de los 21 años) o si existen síntomas atípicos.

### **2.2.3. Diagnóstico diferencial de los trastornos depresivos**

Respecto del **diagnóstico diferencial entre distimia y depresión mayor**, el DSM-IV señala que “el trastorno distímico y el trastorno depresivo mayor se distinguen en base a la gravedad, la cronicidad y la persistencia” (APA, 2002, p. 350). El diagnóstico diferencial entre estos dos trastornos es especialmente complicado debido a que comparten un gran porcentaje de síntomas y porque los criterios temporales establecidos por el manual para distinguirlos son difíciles de valorar. El DSM-IV señala que (APA, 2002, p. 350):

“Si en el inicio de la alteración los síntomas depresivos crónicos son de suficiente gravedad y número como para cumplir los criterios para un episodio depresivo mayor, el diagnóstico debe ser trastorno depresivo mayor, crónico (si todavía se cumplen los criterios), o trastorno depresivo mayor, en remisión parcial (si ya no se cumplen los criterios)”.

En este sentido, para descartar que una persona que cumple los criterios de un trastorno distímico no le corresponda en realidad un diagnóstico de trastorno depresivo mayor en remisión parcial, el evaluador debe comprobar si durante los dos primeros años desde el inicio de la distimia ha cumplido los criterios de episodio depresivo mayor. Este criterio diferencial es, a nuestro modo de ver, no solamente arbitrario, sino poco fiable dada la dificultad de la tarea diagnóstica y la distancia temporal a la que el paciente debe remontarse para dar esta información.

El **trastorno depresivo no especificado**, es un diagnóstico reservado para ser asignado cuando está presente la sintomatología depresiva, pero no se cumplen los criterios de trastorno depresivo mayor, distimia o trastorno adaptativo. En ocasiones está asociado a personas que cumplen criterios de un trastorno de ansiedad. Los principales ejemplos de esta categoría que aporta el DSM-IV son:

- El **trastorno disfórico premenstrual**, caracterizado por una intensa sintomatología (p. ej., ánimo deprimido, ansiedad, labilidad afectiva o pérdida de interés por las actividades). Debe existir al menos una semana libre de síntomas después de las menstruaciones.
- El **trastorno depresivo menor** recoge aquella sintomatología depresiva que durante las dos últimas semanas no ha cumplido los 5 síntomas requeridos en el trastorno depresivo mayor. Por tanto, de entre los 9 síntomas del criterio A del episodio depresivo mayor, se requiere cumplir un máximo de 4. Además, uno de ellos deben ser bajo estado de ánimo o falta de interés/placer. El paciente no ha debido padecer alguna vez un trastorno depresivo mayor y se aplican así mismo los mismos criterios de exclusión de este último trastorno.
- **Trastorno depresivo breve recidivante**: episodios depresivos con una duración de 2 días a 2 semanas, que se presentan al menos una vez al mes durante 12 meses.

### **Diagnóstico diferencial con los trastornos adaptativos**

Los trastornos adaptativos se caracterizan por una reacción emocional o comportamental ante un evento estresante que dura más de tres meses desde el inicio del evento. Una vez terminado el evento estresante, así como sus consecuencias, la reacción no debe durar más allá de 6 meses. La reacción consiste en un “acusado malestar, superior al esperable dada la naturaleza del

estresante, o en un deterioro significativo de la actividad social o profesional (o académica)” (APA, 2002, p. 639). El sujeto no debe cumplir criterios para el diagnóstico de otro trastorno del Eje I o la reacción no se explica completamente por la presencia de éste. Es decir, si el sujeto cumple los criterios del episodio depresivo mayor, no se le asignará el diagnóstico de trastorno adaptativo.

El DSM-IV señala que los trastornos adaptativos son una categoría residual, con una frecuencia de aparición relativamente alta en centros ambulatorios (5-20%), aunque no señala ningún estudio de referencia que apoye estas cifras.

Algunos autores utilizan un criterio de validez para proponer la eliminación del trastorno adaptativo porque “diferenciar, por ejemplo, el trastorno adaptativo con el estado de ánimo deprimido del trastorno depresivo menor, es inherentemente subjetivo y no fiable. Otra cuestión, todavía más importante, la constituye el hecho de que hoy en día se asume en la nosología actual que el trastorno adaptativo y la depresión menor son dos cosas distintas” (Pincus et al., 2005).

### **Diagnóstico diferencial con los trastornos de ansiedad**

Como ya se ha señalado, los trastornos depresivos y los trastornos de ansiedad tienen una alta comorbilidad y la sintomatología subumbral coexiste en un alto porcentaje de pacientes. Este hecho dificulta el diagnóstico diferencial. Por ejemplo, es frecuente que los pacientes con sintomatología depresiva tengan tanto preocupaciones recurrentes como crisis de ansiedad y a su vez, los pacientes con sintomatología de ansiedad, manifiestan con frecuencia bajo estado de ánimo. Por otro lado, asumiendo que en un alto porcentaje de pacientes ambos trastornos tengan una misma etiología, probablemente el tratamiento que se ponga en marcha para uno de los trastornos será eficaz para ambos.

Especialmente complicado es el diagnóstico diferencial entre el trastorno de ansiedad generalizada y el trastorno depresivo mayor. Ambos trastornos comparten sintomatología como las alteraciones del sueño, la fatiga, la irritabilidad, la inquietud o las dificultades de concentración. Además, en el caso del trastorno depresivo mayor, aparecen con una alta frecuencia pensamientos rumiativos que no son fáciles de diferenciar de las preocupaciones excesivas presentes en el trastorno de ansiedad generalizada.

### **Diagnóstico diferencial con las demencias**

Para la valoración de este diagnóstico diferencial se debe tener en cuenta el estado general de salud del paciente y la evolución del deterioro cognitivo en relación con la sintomatología depresiva. Según el DSM-IV (APA, 2002, p. 419):

“En la demencia suele haber una historia premórbida de deterioro de las funciones cognoscitivas, mientras que el sujeto con un trastorno depresivo mayor es mucho más probable que tenga un estado premórbido relativamente normal y un deterioro cognoscitivo brusco asociado a la depresión”.

Según Lishman (1998, p. 150), este complejo análisis debe tener en cuenta que:

“Las características que pueden ayudar a distinguir la depresión primaria de un trastorno psiquiátrico orgánico incluye la evaluación cuidadosa del establecimiento de los fallos/errores de concentración y memoria. En la depresión puede ser observado que la pérdida de interés o ansiedad puede hacer que se pierda el foco de atención, más que ninguna dificultad profunda en la organización del pensamiento”.

**Capítulo 3.**

**INSTRUMENTOS**

**PARA LA**

**EVALUACIÓN Y EL**

**DIAGNÓSTICO DE**

**LOS TRASTORNOS**

**DEPRESIVOS**



En los dos primeros capítulos se ha definido la naturaleza de los trastornos depresivos, su importancia sociosanitaria y las dificultades que existen para su clasificación y diagnóstico. En este tercer capítulo se describirán los instrumentos disponibles para su evaluación y diagnóstico.

Como se ha comentado, hasta el momento la neurociencia no ha desarrollado métodos para el diagnóstico de los trastornos depresivos y tampoco se han encontrado marcadores genéticos para su detección ni indicadores causales suficientes de cualquier tipo (psicológico, social o biológico). El procedimiento más utilizado para la evaluación y el diagnóstico de los trastornos mentales son las entrevistas no estructuradas o clínicas. En este tipo de entrevistas cada profesional selecciona de una manera libre las preguntas a realizar, partiendo de sus propias habilidades y conocimientos, siempre en interacción con el paciente. Aunque la mayoría de las evaluaciones clínicas se siguen realizando con métodos informales de búsqueda de información, muchas organizaciones profesionales y científicas tratan de mejorar la fiabilidad en el diagnóstico mediante guías que orienten la evaluación (Feliciano y Gum, 2010). Los cuestionarios, inventarios o escalas de autoinforme son recomendados porque ayudan en muchos aspectos en la recogida de información sobre el paciente (p. ej., detección de los trastornos, cuantificación de la gravedad de los síntomas, medición de la eficacia de los tratamientos,...) y, aunque hay dudas sobre su utilidad para llevar a cabo diagnósticos, algunos de ellos se han desarrollado precisamente para este objetivo.

La selección de los instrumentos adecuados para la evaluación de los trastornos depresivos juega un papel importante en la psicología clínica. La comunidad científica y los profesionales de la salud mental otorgan un papel cada vez más importante a la asistencia empíricamente validada, necesaria para aumentar la eficiencia de los recursos y proporcionar un adecuado tratamiento a la



población (Hunsley y Mash, 2011), especialmente en el caso de los trastornos depresivos (Pignone et al., 2002).

### **3.1. Entrevistas clínicas**

Las entrevistas no estructuradas, libres o clínicas son el instrumento de evaluación más flexible y adaptado al paciente, de manera que sus resultados están influidos por múltiples aspectos relacionados tanto con el entrevistador (p. ej., sus intereses profesionales, sus intuiciones, su grado de formación, la corriente teórica a la que pertenece o sus impresiones sobre el paciente e incluso por su estado de ánimo), como con el tipo de respuesta del paciente a sus preguntas. “Las formulaciones idiosincrásicas de las hipótesis iniciales, los modos individuales de preguntar sobre los síntomas o la experiencia de enfermedad y el modo individual de valorar el peso de la información obtenida”, influyen en el resultado de las entrevistas (Andrews y Peters, 1998, p. 80). Es por tanto lógico que la variabilidad de estas entrevistas sea elevada y que exista una gran inconsistencia entre diferentes clínicos (Segal, Coolidge, O’Riley y Heinz, 2006).

Para la valoración de los trastornos depresivos en el contexto de la práctica clínica, es fundamental tener en cuenta que el propio bajo estado de ánimo del paciente condiciona el ritmo de la evaluación y sesga sus respuestas. El evaluador debe tener especial cuidado de no interrumpir al paciente y crear un clima de empatía con él. Además, se debe valorar las condiciones y hábitos diarios con el objetivo de establecer su grado de deterioro funcional (Dozois y Dobson, 2011).

La descripción cualitativa de los síntomas es necesaria para conocer los detalles personales del paciente, mejorar su adherencia y personalizar tanto el tratamiento como la evaluación de los cambios terapéuticos. Para conocer con detalle estos síntomas, se deben realizar preguntas concretas acerca de su

frecuencia, intensidad y duración, teniendo en cuenta que los sesgos negativos del paciente pueden dificultar un análisis fiable de su estado. Además, se debe evaluar la personalidad previa del paciente, así como los episodios depresivos anteriores, su duración y el grado de recuperación interepisódico, aspectos fundamentales para el pronóstico del paciente. Las habilidades generales del entrevistador en salud mental son especialmente necesarias para la evaluación de los trastornos depresivos y están descritas con más detalle en otras publicaciones —véase, por ejemplo, Segal, June y Marty (2010) para consultar las habilidades del terapeuta durante las entrevistas y Norcross y Wampold (2011) para las habilidades generales del terapeuta—.

Se ha encontrado que los médicos de atención primaria no diagnostican correctamente el trastorno depresivo mayor (Mitchell, Rao y Vaze, 2011; Thompson, Ostler, Peveler, Baker y Kinmonth, 2001). Fernández et al. (2010) encontraron datos en esta línea en un estudio realizado en Cataluña. Los sanitarios detectaron solamente un 22% de los pacientes depresivos y de los que habían sido señalados como depresivos, un cuarto no cumplía los criterios, con un kappa = 0.16. Para realizar este estudio se comparó el diagnóstico de los médicos con la entrevista SCID-I. Se ha de tener en cuenta que los médicos realizaban el diagnóstico en entrevistas de 7-10 minutos de duración, utilizando códigos CIE-10 (mientras que la entrevista SCID-I sigue los criterios diagnósticos del DSM-IV). Aunque con tasas de reconocimiento más elevadas, sobretudo en pacientes con sintomatología grave, se ha confirmado en otros países la idea de que en atención primaria el grado de acuerdo de los sanitarios con el diagnóstico de las entrevistas estructuradas es bajo (Wittchen y Pittrow, 2002).

Algunos autores señalan que la causa de esta baja fiabilidad en la detección de los trastornos depresivos es precisamente que las formulaciones del DSM-IV

pueden no ser válidas para la clínica diaria, mientras que otros señalan que los médicos son los que deberían evaluar mejor los trastornos depresivos para poder proporcionar a los pacientes que los sufren el tratamiento farmacológico correspondiente (Kessler et al., 2007).

En la práctica diaria tampoco los especialistas en salud mental detectan correctamente los pacientes con trastornos depresivos durante las entrevistas clínicas o no estructuradas. Los índices de fiabilidad diagnóstica continúan siendo más bajos de lo deseable cuando no se utilizan entrevistas estructuradas, como se ha reflejado, por ejemplo, en los estudios de campo del DSM-5 (Regier et al., 2013), con un índice kappa para el diagnóstico del trastorno depresivo mayor de 0.28.

En la Clínica Universitaria de Psicología de la Universidad Complutense de Madrid, donde se llevó a cabo la presente investigación, los terapeutas utilizan como referencia la Guía de Entrevista Clínica General de Muñoz (2003), con el fin de mejorar la fiabilidad y validez de las entrevistas de evaluación. El autor de esta entrevista señala que, además de la información descriptiva y el análisis funcional de las conductas problema del sujeto, se debe realizar un proceso de valoración diagnóstica. Estos tres aspectos están en parte relacionados puesto que los dos primeros proporcionan información necesaria para asignar el diagnóstico y a su vez los diagnósticos proporcionan modelos sintomatológicos y de funcionamiento de los pacientes que ayudan a proseguir con la evaluación, aunque en parte estos procesos también son independientes.

Para evitar confusiones terminológicas, en la presente tesis doctoral se van a considerar *entrevistas estructuradas* y *semiestructuradas* a aquellas en las que los entrevistadores deben leer palabra por palabra las preguntas, sin variar ni su orden ni su contenido. La diferencia principal entre estos dos tipos de entrevistas es que las *semiestructuradas* dejan mayor libertad al clínico para solicitar aclaraciones

o para ampliar las preguntas cuando las respuestas son confusas. Por otro lado, se van a considerar *entrevistas no estructuradas* a aquellas en las que los clínicos tienen total libertad para seleccionar las preguntas y el orden en el que las formulan. Dentro de las *entrevistas no estructuradas* existen igualmente varios grados de estructuración, siendo la entrevista desarrollada por Muñoz (2003) una entrevista flexible que proporciona una cierta guía estructurada (de ahí que se la describa a veces como “semiestructurada”, a pesar de que en esta tesis doctoral se la considera *no estructurada*).

### **3.2. Entrevistas diagnósticas estructuradas y semiestructuradas**

Como ya se ha comentado en capítulos anteriores, durante los años 70 del siglo XX surgió una corriente neokraepeliana que defendía el desarrollo de criterios diagnósticos objetivables, basados en la evidencia y con alta fiabilidad. Las entrevistas estructuradas y semiestructuradas se convierten entonces en una herramienta para el desarrollo de los sistemas diagnósticos, elementos clave en el estudio de su validez y el “patrón oro” para realizar los diagnósticos de los trastornos mentales.

Las entrevistas estructuradas y semi-estructuradas están formadas por una lista estandarizada de preguntas, que se deben realizar sistemáticamente en una secuencia dada y que han de ser valoradas de manera protocolizada a través de un algoritmo de decisión para así llegar al diagnóstico. Estos instrumentos están diseñados para homogeneizar las fuentes de información durante la realización del proceso diagnóstico. Además, sirven para incrementar la cobertura de trastornos, mejorar las habilidades del evaluador para determinar si un síntoma está ausente o presente y reducir la variabilidad entre los entrevistadores (Segal et al., 2006).

Como ya se ha señalado, las entrevistas estructuradas deben ser leídas palabra por palabra, así como las aclaraciones a las preguntas, mientras que en las

semiestructuradas las preguntas de seguimiento pueden ser elaboradas con más libertad, “por ejemplo, el entrevistador puede modificar preguntas existentes en cualquier sentido, e incluso idear unas nuevas y creativas preguntas para valorar más exactamente un síntoma específico” (Segal et al., 2006, p. 122-123). En general, las entrevistas estructuradas pueden ser aplicadas por personal no clínico entrenado, mientras que las entrevistas semiestructuradas deben ser llevadas a cabo por clínicos entrenados, que tengan las habilidades y el juicio clínico necesario. Ambos tipos de entrevistas se utilizan en investigación, en la práctica clínica y en la formación de profesionales. En investigación son utilizadas con frecuencia para asignar a sujetos a condiciones experimentales: “Ciertamente, una buena investigación requiere que individuos asignados a un diagnóstico realmente cumplan todos los criterios para ese diagnóstico” (Segal, et al., 2006, p. 123). Además, estas entrevistas ayudan a valorar si el hecho de que un paciente reciba a lo largo de su vida diagnósticos diferentes se debe a cambios reales en el paciente y no a modificaciones en los terapeutas que le han evaluado o en los métodos de medida que se han utilizado.

Como se puede observar en la Tabla 3.1, existen también desventajas importantes derivadas del uso de las entrevistas estructuradas y semiestructuradas. Al estar centradas exclusivamente en áreas problemáticas del paciente y al requerir que el terapeuta se centre en el seguimiento de las pautas de la entrevista y en la sintomatología del paciente, se puede generar una peor relación terapéutica (Rogers, 2001; Robinson y Asnis, 1989; Segal y Coolidge, 2003, en Segal, et al., 2006), aunque se ha señalado que este problema puede ser resuelto mediante el entrenamiento de los evaluadores (Rogers, 2001, en Segal et al., 2006).

**Tabla 3.1.** Ventajas y desventajas de las entrevistas estructuradas y semiestructuradas (en Segal et al., 2006, p. 125)

Ventajas	Desventajas
<b>Incremento de la fiabilidad:</b> <ul style="list-style-type: none"> <li>- Mejora de la fiabilidad interjueces al disminuir la variabilidad mediante preguntas estandarizadas</li> <li>- Mejora de la fiabilidad en la evaluación del síntoma a través del tiempo, así como de la información proporcionada por el cliente y otras informaciones complementarias</li> </ul>	<b>Puede dificultar la relación terapéutica:</b> porque son entrevistas centradas en el problema, no centradas en la persona, y los evaluadores poco entrenados pueden descuidar las habilidades clínicas básicas durante la evaluación.
<b>Incremento de la validez:</b> debido a que asegura que se evaluarán todos los criterios diagnósticos de manera sistemática.	<b>Están limitadas por el sistema de clasificación en sí mismo:</b> Son válidas solamente en la medida en que son válidos los sistemas sobre los que se han desarrollado. De hecho, la dificultad en la valoración de su validez es que son “el patrón oro” del diagnóstico.
<b>Utilidad como elementos de entrenamiento:</b> <ul style="list-style-type: none"> <li>- Para personal no clínico, que podrá realizar diagnósticos</li> <li>- Para el personal clínico, como elemento de formación en diagnósticos específicos</li> </ul>	<b>Cobertura frente a profundidad:</b> no pueden valorar todas las categorías. Cuando se elige una entrevista estructuradas se debe valorar tanto qué diagnósticos valora como la profundidad con la que lo hace.

### Fiabilidad de las entrevistas estructuradas y semiestructuradas

En el campo de las entrevistas clínicas, este concepto se aplica a dos aspectos diferenciados: la replicabilidad (la estabilidad temporal de los diagnósticos que se realizan mediante las entrevistas) y al grado de coincidencia de los juicios clínicos de diferentes entrevistadores. Por tanto, las dos maneras más habituales de medir la fiabilidad son (Dozois y Dobson, 2011):

- Se comparan los diagnósticos asignados al paciente por parte de distintos evaluadores, a partir de una misma entrevista (fiabilidad interjueces).
- Se realizan dos entrevistas, habitualmente separadas por dos semanas, aplicadas ya sea por el mismo o por entrevistadores distintos (fiabilidad test-retest). Esta alternativa se elige en menos ocasiones. Existe evidencia de que los sujetos suelen responder a ambas entrevistas de manera diferente, encontrándose por ejemplo una mayor probabilidad de respuesta negativa ante las preguntas en la aplicación del retest. Por otro lado, a menudo existen confusiones respecto al diseño metodológico del retest puesto que algunos autores incluyen en él nuevas preguntas con el objetivo de ahondar más en el proceso de evaluación, lo que ha sido considerado como una prueba de validez diagnóstica y no, o no sólo, de fiabilidad test-retest.

La fiabilidad viene influida por el grado de entendimiento del entrevistado, así como por el grado de entrenamiento de los entrevistadores, la prevalencia del trastorno en una población determinada y, en definitiva, por las condiciones de aplicación.

#### **Validez de las entrevistas estructuradas y semiestructuradas**

En la medida en la que las categorías diagnósticas de un sistema diagnóstico no sean válidas, los instrumentos desarrollados a partir de dichos sistemas diagnósticos tampoco lo serán. A falta de un criterio externo diagnóstico con el que ser comparadas como, por ejemplo, los resultados de un prueba de neuroimagen en el caso de que el modelo biomédico de los trastornos mentales fuera cierto, las entrevistas diagnósticas se consideran el “patrón oro” del diagnóstico y, por tanto, son las propias entrevistas las que se validan en relación

con otras. Especialmente en los trastornos depresivos, la entrevista SCID-I-IV (First et al., 1999) es una de las referencias principales.

Como alternativa para alcanzar una mayor validez en el diagnóstico, la iniciativa LEAD (observaciones longitudinales realizadas por expertos usando toda la información disponible) (Spitzer, 1983), también conocido como el *diagnóstico mejor estimado*, persigue aumentar la validez de los diagnósticos recogiendo el máximo de información posible sobre el paciente por parte de varios clínicos. Se utiliza para ello entrevistas, cuestionarios y otros informadores, y el diagnóstico alcanzado sobre la base de toda esa información es considerado el criterio para validar las entrevistas diagnósticas.

En general, se considera que “la validez es presumiblemente mayor en las entrevistas clínicas semiestructuradas que en aquellas entrevistas estructuradas como la CIDI”, aunque las entrevistas estructuradas son consideradas como más fiables (Kessler y Üstün, 2004, p. 99).

A continuación se describirán las principales entrevistas diagnósticas estructuradas y semiestructuradas que incluyen la evaluación y diagnóstico de los trastornos depresivos y cuentan con suficiente apoyo empírico para su uso por los servicios de salud mental y atención primaria. En la Tabla 3.2. se pueden observar las características fundamentales de tales entrevistas.



**Tabla 3.2.** Comparación de las características de las entrevistas diagnósticas (modificado de Summerfeldt, Klossterman y Antony, 2011)

Entrevista	Cobertura	Duración entrevista <sup>º</sup>	Población (objetivo)	D. V.	Evaluadores recomendados	Escalas dimensionales de gravedad	Grado de estructuración	Manual de referencia	Adecuada como diagnóstico <sup>ª</sup>
SCID-I	Superior	45-90 minutos	Población clínica y comunitaria	Sí	Profesionales salud mental entrenados	No (aunque incluye sub-umbrales)	Semiestructurada	DSM-IV	Sí
CIDI	Superior	120 minutos	Población comunitaria	Sí	No profesionales entrenados	Sí, en la versión 3.0	Estructurada	DSM-IV y CIE-10	No
SADS	Media	60 (no pacientes) 120-240 (pacientes)	Población clínica y comunitaria	Sí	Profesionales salud mental entrenados	Sí (menos para psicóticos)	Semiestructurada	Criterios RDC (Versión DSM-IV)	Sí
MINI	Media-Superior	15 minutos	Población comunitaria	Algunos	Versión clínicos y no clínicos entrenados	No	Estructurada	DSM-IV y CIE-10	Sí
PRIME-MD	Media	10-20 minutos	Atención primaria	No	Profesionales de la salud entrenados	No	Semiestructurada	Casi DSM-IV	No
DIS	Media	90-120 minutos	Comunitaria	Sí	No profesionales entrenados	No	Estructurada	DSM-IV	No
SCAN	Superior	90 minutos	Población clínica		Profesionales salud mental entrenados	Sí	Semiestructurada	DSM-IV y CIE-10	Sí

<sup>º</sup>Muestra no psiquiátrica; D.V.: Diagnósticos vitales; <sup>ª</sup> Adecuada como medida principal de diagnóstico psiquiátrico.

- **Entrevista clínica estructurada para los trastornos del EJE I del DSM-IV (SCID-I; *Structured Clinical Interview for DSM-IV*)**

La SCID-I es una entrevista semiestructurada de evaluación de los trastornos mentales que se ha desarrollado, desde su nacimiento en 1983, en paralelo a los cambios realizados en el DSM. Es probablemente la entrevista más utilizada en investigación en EE. UU. De hecho “la entrevista SCID en varios de sus formatos, ha sido utilizada en más de 1000 estudios publicados” (Summerfeldt, Antony y Klossterman, 2011, p. 124). Es la entrevista con más apoyo empírico para el diagnóstico de los trastornos del estado de ánimo (Joiner et al., 2005).

En sus versiones más recientes, permite realizar el diagnóstico de los siguientes trastornos del Eje I del DSM-IV: trastornos del estado de ánimo, trastornos de ansiedad, esquizofrenia, trastornos somatomorfos, trastornos adaptativos, trastornos por dependencia de sustancias y trastornos de alimentación, aunque recientemente se ha añadido un módulo de trastornos del control de impulsos. Hay dos versiones disponibles, la SCID-I-RV (versión para investigación) (First, Spitzer, Gibbon y Williams, 2002) y la SCID-I-CV (versión clínica), que no incluye los trastornos somatomorfos ni los trastornos de alimentación (First, Spitzer, Gibbon y Williams, 1997). La SCID-I-RV incluye también las especificaciones correspondientes a los diagnósticos DSM-IV, que en el caso de la depresión mayor son: con patrón estacional, con inicio en el posparto, con síntomas melancólicos, con síntomas atípicos y con síntomas catatónicos; mientras que en la versión SCID-I-CV estas especificaciones no están presentes.

La versión clínica de la SCID-I puede ser aplicada tanto a población psiquiátrica como a población general en personas mayores de 18 años (aunque con pequeñas modificaciones también se puede aplicar a adolescentes).

Debe ser aplicada por expertos clínicos entrenados, debido a que es una entrevista semiestructurada que necesita del juicio clínico para la valoración de los criterios y permite añadir comentarios o repreguntas para aclarar las respuestas de los sujetos. Persigue la valoración tanto de los trastornos actuales como de los ocurridos durante la vida del paciente (en el caso de los trastornos depresivos). Se especifica en su manual que se puede aplicar la entrevista completa o módulos específicos (como es el caso del módulo de los trastornos del estado de ánimo). Esto es de gran utilidad porque el evaluador puede personalizar la entrevista según las necesidades de cada paciente (Segal et al., 2006). Según el manual de usuario de la SCID-I, se puede utilizar esta entrevista como confirmación de un diagnóstico que el clínico sospecha, aplicar la entrevista completa para una mayor cobertura diagnóstica o administrarla para la formación de especialistas en salud mental con el objetivo de mejorar sus habilidades y conocimientos sobre los criterios diagnósticos (First et al., 1999).

La entrevista comienza con un módulo con preguntas abiertas, que valora información sociodemográfica, estatus laboral, historia psicopatológica previa y de tratamientos anteriores y funcionamiento actual. Los autores de la entrevista señalan que esto prepara la relación con el paciente para las preguntas más estructuradas y proporciona al evaluador un contexto de interpretación de las preguntas diagnósticas (First et al., 1999).

El investigador leerá las preguntas palabra por palabra y debe pedir aclaraciones respecto a aquellas respuestas que simplemente los pacientes respondan con un “sí” o “no” poco elaborados. El árbol de decisión de la entrevista sigue los criterios del DSM-IV, de manera que cuando el terapeuta descarta la presencia de un criterio necesario para el diagnóstico, se salta el resto de

preguntas, lo que provoca que se pierda la información sobre la sintomatología subclínica.

El aplicador de esta entrevista tendrá los criterios DSM-IV siempre presentes, además del texto a partir del cual debe realizar las preguntas. Para valorar cada criterio existen cuatro opciones de codificación de la respuesta del sujeto: el criterio puede ser positivo en caso de presencia del síntoma (+), subclínico (en el caso en caso de que esté presente pero no cumpla con la duración o la intensidad requerida) o ausencia del síntoma (-), e información insuficiente (?), en el caso de que el evaluador requiera información adicional para formarse un juicio clínico sobre su presencia.

#### **Fiabilidad**

Se ha comentado que dentro del conjunto de las entrevistas diagnósticas, la SCID-I tiene estadísticos de fiabilidad moderados, aunque los estudios son por lo general de versiones anteriores de la entrevista (Dozois y Dobson, 2011). En la práctica real los aplicadores no están entrenados como en las investigaciones, lo que puede provocar menor fiabilidad en el diagnóstico. Por tanto, es un instrumento especialmente sensible al entrenamiento y la habilidad del clínico que la aplica.

Summerfeldt, et al. (2011, p. 126) señalan que en “En general, se han encontrado aceptables cifras de fiabilidad”, ( $\kappa$  por encima de 0.75) en la mayoría de los estudios de los diagnósticos más comunes en muestras clínicas, como el trastorno depresivo mayor y los trastornos de ansiedad.

En la página web oficial de la entrevista SCID (<http://www.scid4.org>; *Biometrics Research Department at Columbia University*) se puede encontrar una revisión de los estudios de fiabilidad (véase lea Tabla 3.3) y, efectivamente, tan solo se ha publicado un estudio significativo sobre su fiabilidad durante la última década.

**Tabla 3.3.** Características e índices de fiabilidad (estadístico kappa) para los trastornos depresivos de los estudios sobre la fiabilidad de la entrevista SCID-I (*Biometrics Research Department at Columbia University*; <http://www.scid4.org>)

	Skre et al., 1991	Zanarini et al., 2000	Zanarini et al., 2000	Segal, et al., 1995	Williams et al., 1992	Zanarini et al., 2001	Zanarini et al., 2001	Lobbetael et al., 2010
Muestra	N = 54	N = 27	N = 52	N = 40	N = 592	N = 45	N = 30	N = 151
Versión de SCID-I	DSM-III-R	DSM-IV	DSM-IV	DSM-III-R	DSM-III-R	DSM-III-R	DSM-III-R	DSM-IV
TDM	0.93	0.80	0.61	0.90	0.64	0.90	0.73	0.66
Distimia	0.88	0.76	0.35	0.53	0.40	0.91	0.60	0.81

*Nota* TDM = Trastorno depresivo mayor.

Aunque en general los datos de fiabilidad son aceptables, el tamaño de las muestras es escaso en la mayoría de los estudios. En la investigación con mayor tamaño muestral realizada según los criterios DSM-IV (Lobbetael, Leurgans y Arnt, 2011), la fiabilidad del diagnóstico del trastorno depresivo mayor fue de kappa igual a 0.66, lo que se podría calificar como una fiabilidad buena (Altman, 1991).

### Validez

El contenido de la última versión de la entrevista concuerda en la práctica con los criterios del DSM-IV. Las preguntas están directamente relacionadas con los criterios diagnósticos, que deben estar en todo momento presentes para los entrevistadores. Existen pocos datos acerca del grado de acuerdo de la entrevista con evaluaciones clínicas u otros instrumentos diagnósticos, debido a que existen dificultades para encontrar una técnica de evaluación con la que pueda ser comparada (First et al., 1999). La SCID-I es considerada el “patrón oro” del diagnóstico de los trastornos mentales, así que los estudios están dirigidos precisamente a validar otros instrumentos con respecto a esta entrevista. Se ha

encontrado desde una buena a una aceptable concordancia entre los diagnósticos realizados con la entrevista CIDI en sus primeras versiones con respecto a los diagnósticos realizados por la SCID-I en el caso de los diagnósticos DSM-III (Wittchen, 1994; Wittchen et al., 1995; Wittchen et al., 1996, y Kessler et al., 1998, en Haro et al., 2006b).

En el caso de los estudios realizados con la CIDI 3.0 los datos son muy escasos. Dentro del estudio realizado en EE. UU. denominado *National Comorbidity Survey Replication* (NCS-R), después de la investigación epidemiológica llevada a cabo mediante la entrevista CIDI, se aplicó la entrevista SCID-I a un subgrupo de la muestra total:

“El estudio de replicación fue diseñado para determinar el grado en el que las clasificaciones diagnósticas realizadas sobre la base de la CIDI habrían sido diferentes si el estudio hubiese sido realizado por entrevistadores clínicos cuidadosamente entrenados usando la SCID-I en lugar de entrevistadores no especialistas usando la CIDI” (Haro et al., 2006b, p. 170).

Los resultados acerca de la prevalencia vital encontrada a través de ambas entrevistas indicaron un grado de acuerdo dispar y una concordancia entre aceptable y excelente (véase la Tabla 3.4.). En concreto, respecto al trastorno depresivo mayor, se encontró que entre ambas entrevistas había una concordancia moderada,  $\kappa = 0.54$ . El área bajo la curva ROC fue moderado ( $AUC = 0.75$ ), mientras que la sensibilidad fue de 55.3%, lo que indicaba que ese mismo porcentaje de sujetos diagnosticados como trastorno depresivo mayor mediante la entrevista SCID-I también fueron detectados del mismo modo mediante la entrevista CIDI. Mejores fueron los datos sobre el poder predictivo positivo (73.3%), el poder predictivo negativo (86.8%) y la especificidad (93.7%) (Haro et al., 2006b).

**Tabla 3.4.** Consistencia del diagnóstico vital DSM-IV entre la entrevista CIDI y la SCID-I en la reevaluación de una muestra del estudio NCS-R ( $N = 325$ ) (Haro et al., 2006b)

	<b>AUC</b>	<b>Kappa</b>	<b>Sensibilidad</b>	<b>Especificidad</b>	<b>PPP</b>	<b>PPN</b>
TDM	0.75	0.54	55.3	93.7	73.7	86.8
Trastornos de ansiedad	0.73	0.48	54.4	90.7	74.5	80.0
Cualquier trastorno	0.76	0.52	62.8	89.9	84.3	71.7

*Nota.* TDM = trastorno depresivo mayor. AUC = área bajo la curva ROC (*area under curve*). PPP = valor predictivo positivo. PPN = valor predictivo negativo.

Para el estudio de la prevalencia en los últimos 12 meses, el diagnóstico de la SCID-I se realizó 2 meses después de la aplicación de la CIDI sobre una muestra del estudio ESEMeD (véase la Tabla 3.5.). Los datos señalaron una buena eficacia diagnóstica, con áreas bajo la curva ROC en un rango entre 0.8 y 0.9 para los trastornos depresivos, el conjunto de los trastornos de ansiedad y el conjunto de los trastornos metales. Para el conjunto de los trastornos depresivos, existía una gran consistencia entre países con un área bajo la curva ROC entre 0.83 y 0.84. La concordancia volvió a ser en el caso de los trastornos depresivos aceptable, kappa = 0.56, y, en general, los datos de eficacia diagnóstica fueron positivos (área de la curva ROC = 0.83). En este caso cabe destacar que una mayoría de pacientes diagnosticados por la SCID-I con un trastorno depresivo, fue detectado también por la CIDI (69.1%). Los autores señalaban que:

“Hemos observado que los umbrales diagnósticos para la prevalencia vital de la CIDI para los trastornos del DSM-IV son generalmente de alguna manera más conservadores que los de la SCID-I, al menos en Estados Unidos, mientras que los umbrales para la prevalencia a los doce meses generalmente no tienen sesgos, al menos en los 3 países analizados aquí (Francia, España e Italia)” (Haro et al., 2006b, p. 177).

**Tabla 3.5.** Consistencia del diagnóstico CIDI y SCID-I según el diagnóstico DSM-IV en los últimos 12 meses en la reevaluación clínica de una muestra del estudio ESEMeD y con menos de 60 días de diferencia entre las aplicaciones de las dos entrevistas ( $n = 143$ ) (Haro et al., 2006b)

	<b>AUC</b>	<b>Kappa</b>	<b>Sens.</b>	<b>Especificidad</b>	<b>PPP</b>	<b>PPN</b>
TDM	0.83	0.56	69.1	97.2	49.6	98.7
Trastornos de ansiedad	0.88	0.42	83.7	92.9	31.3	99.3
Cualquier trastorno	0.84	0.49	77.9	90.6	41.5	98.0

*Nota.* TDM: trastorno depresivo mayor; Sens: sensibilidad; AUC: área bajo la curva ROC; PPP: poder predictivo positivo; PPN: poder predictivo negativo.

Además, el hecho de que la SCID-I tenga una fiabilidad test-retest muy alejada de la perfección (Segal et al., 2006), especialmente en muestras epidemiológicas (Williams et al., 1992), y el hecho de que algunos sujetos esconden información en los estudios epidemiológicos (Kranzler et al., 1997, en Haro et al., 2006b), pueden hacer que en ese estudio se estuviera infravalorando la validez de la CIDI. Por otro lado, existe un posible sesgo en estos procesos de validación puesto que los clínicos que debían aplicar la entrevista SCID-I, aunque no conocían las contestaciones de los sujetos en cada criterio, sí conocían el diagnóstico final asignado (Alegría et al., 2009).

### **Versión española**

En el manual publicado en España de la versión clínica de la SCID-I no se informa de los datos de fiabilidad de esta adaptación, tan solo se aportan la información que viene reflejada en la versión original de la entrevista (First et al., 1999).

En una investigación realizada sobre una muestra de 105 sujetos reclutados en un centro de abusos de sustancias se comparó el diagnóstico realizado mediante la versión clínica de la edición española de la SCID-I con el diagnóstico



DSM-IV asignado mediante el procedimiento LEAD. En el caso de los trastornos depresivos, mediante la entrevista SCID se diagnosticó un porcentaje significativamente mayor de personas (41% en el caso del trastorno depresivo mayor) frente al diagnóstico LEAD (28%). El grado de concordancia entre ambos fue bajo en el caso del trastorno depresivo mayor actual ( $\kappa = 0.28$ ) y aceptable en el caso de los trastornos depresivos a lo largo de la vida ( $\kappa = 0.54$ ) (Torrens, Serrano, Astals, Perez-Dominguez y Martin-Santos, 2004).

- **Entrevista Diagnóstica Compuesta Internacional (CIDI; *Composite International Diagnostic Interview*)**

La CIDI es una entrevista completamente estructurada, desarrollada por la Organización Mundial de la Salud para los principales trastornos mentales. En su última versión CIDI 3.0 (Kessler y Üstün, 2004) permite realizar diagnósticos tanto en códigos DSM-IV como CIE-10.

La CIDI es la entrevista estructurada más utilizada en los estudios epidemiológicos en salud mental y permite ser aplicada por personal entrenado no especializado. Es necesario realizar las preguntas de la entrevista siguiendo cada módulo palabra por palabra y en el orden prefijado, de manera que no se debe omitir ninguna pregunta y si se quiere añadir alguna aclaración esta debe ser incluida al final de cada módulo. Permite realizar diagnósticos del estado del paciente durante los últimos 12 meses o a lo largo de la vida del paciente.

Esta entrevista fue desarrollada a partir de la entrevista DIS (Robins, Helzer, Croughan y Ratcliff, 1981), que originalmente solo incluía criterios DSM-III, en conjunción con la entrevista FSE (Wing, Cooper y Sartorius, 1974), desarrollada para valorar el estado mental de los sujetos en el presente. La CIDI fue diseñada para fomentar los estudios epidemiológicos comunitarios en países de todo el

mundo. Para ello, un comité editorial internacional de la CIDI se encarga de realizar estudios de campo en muchos países diferentes (Kessler y Üstün, 2004).

Existen sucesivas versiones de la entrevista CIDI, que progresivamente ha crecido tanto en la cobertura de los trastornos que evalúa como en la inclusión de factores sociodemográficos, factores funcionales, factores de riesgo o la historia de tratamiento. La versión actual CIDI 3.0 tiene una duración aproximada de 2 horas (Kessler y Üstün, 2004).

También existe una versión abreviada para los trastornos depresivos, que tiene unas buenas cualidades psicométricas comparada con la CIDI completa (Kessler, Mroczek, 1994a, en Wang et al., 2010). Esta versión incluye una sección inicial con preguntas de cribado o *screening* y de prevalencia vital, y recoge, entre otros, los diagnósticos de los trastornos del estado de ánimo (depresión y manía).

#### **Fiabilidad**

Los estudios realizados con la entrevista CIDI señalan unos buenos datos de fiabilidad, probablemente debido al alto grado de estructuración de la entrevista (Andrews y Peters, 1998), con kappas por encima de 0.9 en 17 de sus 20 escalas, por lo que estos autores señalan que “la CIDI es en la actualidad la entrevista estructurada más fiable” (Andrews y Peters, 1998, p. 86).

Los estudios de fiabilidad de la CIDI han recibido la crítica de haber sido realizados principalmente en contextos clínicos más que en muestras generales, dónde la prevalencia del trastorno es menor, mientras que este instrumento está diseñado para ser aplicado en grandes muestras de población general, por lo que la fiabilidad puede cambiar de manera significativa. Además, Kurdyak y Gnam (2005, p. 853) afirman que:

“Algunos problemas acompañan las interpretaciones de los estudios de fiabilidad de la CIDI. Un asunto importante es que la CIDI tiene modificaciones significativas desde su creación. Las mediciones de fiabilidad de algunas versiones no tienen que ser necesariamente generalizables a otras. De manera similar, la fiabilidad de algunos estudios están dirigidos a módulos diagnósticos distintos de la depresión, y pueden no reflejar específicamente las propiedades del módulo de depresión”.

Estos autores añaden que “la CIDI... tiene una buena fiabilidad interjueces, aunque en el módulo de depresión en la forma actual no ha sido rigurosamente evaluado” (p. 853).

No se han publicado recientemente estudios de fiabilidad del diagnóstico de la entrevista CIDI. Se han publicado, sin embargo, trabajos que analizan con profundidad las dificultades teóricas y prácticas de este tipo de estudios (Kessler y Üstün, 2004), pero desde los años 90 del siglo pasado:

“Solo un puñado de estudios de validez sobre la CIDI o la DIS han sido publicados... Los resultados muestran que tanto los diagnósticos de la DIS como de la CIDI están significativamente relacionados con diagnósticos clínicos independientes, pero la concordancia individual está lejos de la perfección. Una parte de esta falta de concordancia sin duda se debe a la falta de fiabilidad de las entrevistas clínicas. De hecho, la literatura es clara en mostrar que la fiabilidad test-retest de las clasificaciones diagnósticas basadas en la DIS-CIDI es mayor que con las entrevistas clínicas semiestructuradas” (Kessler y Üstün, 2004, p. 96).

De manera que durante las últimas décadas estos investigadores han centrado su atención en la validez de la entrevista, que “la literatura mostrado claramente que es presumiblemente mayor en las entrevistas semiestructuradas que en las completamente estructuradas” (Kessler y Üstün, 2004, p. 96).

### **Validez**

Ya se han comentado los datos de dos estudios de la CIDI 3.0 con respecto a la entrevista SCID-I, con un resultado de validez diagnóstica que se podría calificar de moderado-alto.

Con respecto a los estudios sobre la concordancia de los diagnósticos realizados con la CIDI y la SCAN, es difícil llegar a conclusiones porque la SCAN por sí misma no ha sido validada adecuadamente, aunque para la depresión, el acuerdo entre los dos instrumentos es débil, con un kappa = 0.39 para el diagnóstico de depresión actual en un estudio (Andrews, Peters, Guzman y Bird, 1995) y con una sensibilidad y una especificidad de 0.5 y 0.9 en otro estudio (Brugha, Jenkis, Taub, Meltzer y Bebbington, 2001). Además, en otro estudio se utilizó la entrevista SCAN como “patrón oro” para determinar la validez de la CIDI 2.1 y se encontraron datos de acuerdo entre débiles y moderados ( $k = 0.31$  para los episodios depresivos), con un porcentaje alto de falsos positivos (Jordanova, Wickramesinghe, Gerada y Prince, 2004).

### **Adaptación española**

La adaptación española de la CIDI 3.0 forma parte de la investigación ESEMeD. Como hemos comentado anteriormente, los resultados del estudio de validez diagnóstica con respecto a la SCID-I indicaban un buen rendimiento diagnóstico, con un índice de concordancia aceptable en el caso de los trastornos del trastorno depresivo mayor (Haro et al., 2006b).

- **Protocolo para los Trastornos Afectivos y la Esquizofrenia (SADS; *Schedule for Affective Disorders and Schizophrenia*)**

El SADS es una entrevista semiestructurada originalmente diseñada para valorar los criterios diagnósticos de investigación (RDC) de Spitzer, Endicott y Robins (1978). Existen muchas versiones de esta entrevista a lo largo de sus más de 30 años de desarrollo. La versión normal (SADS) tiene dos partes, la primera cubre los diagnósticos más actuales y la segunda cubre los diagnósticos pasados. Progresivamente se han generado versiones que tienen en cuenta los criterios DSM, como la versión SAD-LA-DSM-IV (Fyer, Endicott, Mannuza y Klein, 1995, en Dozois y Dobson 2011).

Al estar basado en los RDC, la amplitud de diagnósticos evaluados es menor y cubre, en el caso de los trastornos del estado de ánimo, el trastorno depresivo mayor, el trastorno bipolar y el trastorno depresivo menor. Además incluye 8 escalas dimensionales, entre las que se encuentra una del estado de ánimo deprimido, lo que aporta información clínica muy útil en la evaluación de los trastornos del estado de ánimo.

El SADS realiza diagnósticos del estado actual y del periodo de mayor sintomatología del último episodio, con el objetivo de mejorar la fiabilidad de la evaluación. Se puntúan la mayoría de los elementos en una escala de gravedad de 1 a 6, siendo, por ejemplo, 1 = síntoma no presente en absoluto; 3 = síntoma moderado, una frecuencia de síntoma o una intensidad de baja a media; 6 = síntoma extremo, un síntoma que no remite o de una intensidad alta.

Es claramente una entrevista semiestructurada porque, además de las preguntas estandarizadas, se solicita al investigador que recabe pruebas adicionales y realice preguntas individualizadas para clarificar las respuestas de los

sujetos. En esta entrevista es especialmente importante poner en marcha las habilidades del entrevistador y tener familiaridad con los criterios diagnósticos, por lo que es necesario el entrenamiento por parte del personal clínico.

La duración de la entrevista es de 1 hora con muestra no clínica. En el caso de muestras psiquiátricas su duración es entre 2 y 4 horas.

### **Fiabilidad**

Las propiedades psicométricas del SADS han sido estudiadas con detenimiento, pero las múltiples versiones diferentes de la entrevista hacen difícil resumir sus características (Summerfeldt et al., 2011).

En lo referido a la fiabilidad del instrumento, ha mostrado desde buena a excelente fiabilidad (Rogers, 2001). Los propios creadores del test (Spitzer et al., 1978), en sus estudios test-retest, encontraron concordancias altas ( $\kappa$  medio = 0.91 en trastornos actuales y  $\kappa$  medio = 0.93 en trastornos a lo largo de la vida). Exceptuando estos estudios iniciales, los demás estudios incluidos por Rogers (2001) en su revisión se componían en general de muestras pequeñas y con resultados muy dispares.

### **Validez**

Se han encontrado buenos datos de validez para el SADS (Dozois y Dobson, 2011). Con respecto a los trastornos depresivos, a través del SADS se ha encontrado una buena capacidad para predecir el curso del problema y el resultado de los tratamientos (Coryell et al., 1994). Aunque en algunos trastornos los criterios se alejan del DSM-IV, en realidad los trastornos depresivos casi han permanecido constantes desde la elaboración de los RCD.

### **Adaptación española**

No se podido localizar ninguna adaptación española del SADS.

#### **- Entrevista Neuropsiquiátrica Internacional (MINI; (*Mini-International Neuropsychiatric Interview*))**

Diseñada por un equipo de psiquiatras y médicos en Francia y Estados Unidos con el objetivo de evaluar los trastornos mentales en ensayos clínicos y estudios epidemiológicos y realizar diagnósticos DSM-IV y CIE-10 (Sheehan, et al., 1998). Su tiempo de administración es de 15 minutos y puede ser aplicada por personal no clínico. En la actualidad ha sido traducida a más de 40 idiomas (Summerfeldt et al., 2011).

La entrevista MINI se centra en los síntomas principales de los trastornos y a excepción de la valoración del trastorno bipolar, fija su atención en los síntomas manifestados durante las últimas semanas (2 semanas en el caso de la depresión mayor). Por tanto, se centra exclusivamente en episodios presentes en el momento de la evaluación. Fue creada con el ánimo de ser sensible, aún a costa de ser demasiado inclusiva y asignar diagnósticos psiquiátricos a personas que no les corresponde, “aceptando unos poco falsos positivos antes que perder casos verdaderos” (Sheehan et al., 1998, p. 23).

Esta entrevista cuenta con la posibilidad de ser aplicada por el clínico (MINI-CR), pero también tiene una versión autoaplicada (MINI-PR) (Lecrubier et al., 1997). Además, existe una edición más extendida, la entrevista MINI Plus que incluye 23 trastornos diferentes, historia de trastornos ocurridos a lo largo de la vida, especificaciones para los diagnósticos y cronología de los trastornos. También se han desarrollado la versión MINI-Screen (aproximadamente 5 minutos de aplicación) y la versión para niños y adolescentes MINI-Kid (Sheehan et al., 1998).

Joiner et al. (2005, p. 273) señalan que:

“La SCID tiene la ventaja sobre la MINI de cubrir una mayor variedad de trastornos, un historial más amplio y el subtipo atípico, tener una mayor cobertura del curso y cronicidad, e incorporar las ventajas de las escalas clínicas y de los autoinformes. La MINI, por su parte, es más breve y más fácil de aprender y utilizar, y proporciona una buena cobertura del subtipo melancólico”.

#### **Fiabilidad**

En el estudio de validación de la publicación original de esta entrevista se analizaron los datos de fiabilidad en el diagnóstico. En el caso de la depresión mayor la fiabilidad interjueces fue de 1 y la fiabilidad rest-retest de 0.87. En el caso de la distimia no se incluyeron datos por tener una muestra insuficiente (Lecrubier et al., 1997).

#### **Validez**

En la investigación para la validación del MINI 5.0, se compararon los diagnósticos realizados mediante esta entrevista con los realizados mediante las entrevistas SCID-I y CIDI en dos muestras estadounidenses y francesas respectivamente ( $N = 84$ ), con el manual DSM-III-R como referencia (Sheehan et al., 1997). Los resultados (véase la Tabla 3.6.) señalaron, en el caso del trastorno depresivo mayor, unas buenas cualidades psicométricas, en especial de la MINI-CR con respecto a la SCID ( $\kappa = 0.84$ ). Las sensibilidades, especificidades, valores predictivos positivos y valores predictivos negativos fueron mayores de 0.87 con respecto a la entrevista SCID, mientras que en el caso de la MINI-PR ( $\kappa = 0.55$ ), la sensibilidad y el valor predictivo positivo fueron algo menores (0.77 y 0.75 respectivamente) (Sheehan et al., 1997).



**Tabla 3.6.** Concordancia entre los diagnósticos de trastorno depresivo mayor y distimia de la entrevista MINI-CR (aplicación clínica) y MINI-PR (autoaplicada) con respecto a la SCID-I y la CIDI (Lecrubier et al., 1997, p. 228; Sheehan et al., 1997, p. 235)

Entrevistas	Trastorno	Kappa	Sensibilidad	Especificidad	PPP	PPN
MINI-CR	TDM	0.84	0.96	0.88	0.87	0.97
con SCID	Distimia	0.52	0.67	0.99	0.45	0.99
MINI-PR	TDM	0.55	0.77	0.79	0.75	0.81
con SCID	Distimia	0.11	0.17	0.98	0.11	0.98
MINI-PR con CIDI	TDM	0.73	0.94	0.79	0.82	0.93

*Nota.* TDM: trastorno depresivo mayor; PPP: poder predictivo positivo; PPN: poder predictivo negativo.

Cuando se comparó la MINI con el diagnóstico realizado por la entrevista CIDI, la entrevista MINI-CR mostró igualmente una buena concordancia ( $\kappa = 0.73$ ) y valores de sensibilidad y especificidad de 0.94 y 0.79. Estos valores respaldan el uso de la entrevista MINI para el diagnóstico del trastorno depresivo mayor (Lecrubier et al., 1997).

Los datos respecto a la distimia fueron menos favorables: la entrevista MINI-CR tuvo una concordancia aceptable con respecto a la SCID ( $\kappa = 0.52$ ), pero el poder predictivo positivo fue bajo, mientras que la concordancia fue inaceptable en el caso de la MINI-PR ( $\kappa = 0.11$ ) (Lecrubier et al., 1997).

Existen numerosos estudios sobre la validez diagnóstica de la MINI realizados en otros países (p. ej., Mordal, Gundersen y Bramness, 2010; Rossi et al., 2004), que ofrecen aceptables datos de fiabilidad y validez.

### Adaptación española

En el estudio de validación de esta entrevista para población española se comparó el diagnóstico realizado por médicos de atención primaria mediante

entrevistas clínicas (utilizadas en este caso como “patrón oro”) con la versión española de la MINI. Este estudio se realizó en dos provincias españolas sobre una muestra de 126 personas todas las cuales se evaluaron mediante la adaptación española de la entrevista MINI. Los resultados que se presentan con respecto al trastorno depresivo mayor mostraron una alta sensibilidad (94.1) y una moderada especificidad (62.2), con un poder predictivo negativo bajo (41.0) y un poder predictivo positivo excelente (97.4) (Bobes, 1998).

- **Evaluación de los Trastornos Mentales en Atención Primaria (PRIME-MD; *Primary Care Evaluation of Mental Disorders*)**

Esta breve entrevista semiestructurada tiene como objetivo la detección de los trastornos mentales por parte de personal de atención primaria entrenado para su aplicación. Cubre el trastorno depresivo mayor, la distimia, el trastorno depresivo menor, el trastorno de pánico, otros trastornos de ansiedad, el abuso de alcohol, la somatización y los trastornos de la alimentación (Spitzer et al., 1994).

Consta de dos partes diferentes. En primer lugar, un cuestionario administrado por parte del clínico (*Patient Health Questionnaire* o PHQ), con 25 preguntas de respuesta dicotómica si/no sobre una variedad de síntomas y que se aplica en pocos minutos. Si el paciente cumple los criterios de alguno de los trastornos, se administra la segunda parte, la Guía de Evaluación Clínica (CEG), una entrevista semiestructurada que persigue el diagnóstico del paciente. Su tiempo de aplicación media es de 8 minutos, con menos de 20 minutos en un 95% de las ocasiones (Spitzer et al., 1994).

#### **Fiabilidad**

Se han realizado poco estudios adicionales sobre la fiabilidad de este instrumento. En un estudio realizado en Holanda en una muestra de paciente en

atención primaria, se utilizó la entrevista telefónica PRIME-MD. Se calculó la fiabilidad test-retest, con una semana entre cada aplicación. Se encontró en el caso del trastorno depresivo mayor un grado de acuerdo moderado,  $\kappa = 0.58$ . (Bakker et al., 2009).

#### **Validez**

En el estudio de validación original, se entrevistaron a 3 000 personas, de las que 292 fueron detectados por el PHQ como pacientes con trastorno depresivo mayor y 184 con otro trastorno depresivo (Spitzer, Kroenke, Williams, 1999). Los análisis de eficacia diagnóstica con respecto al diagnóstico realizado por los clínicos señalan una sensibilidad del 73% y una especificidad del 94%. La correlación en la gravedad de los síntomas entre ambos sistemas de evaluación fue de 0.84. El tiempo de revisión del cuestionario por parte del especialista es menor a 3 minutos en el 86% de los casos. Los autores señalan que además de proporcionar un diagnóstico provisional, proporcionan un índice de gravedad de los síntomas con adecuadas cualidades psicométricas.

En otra investigación realizada por los creadores del instrumento, se compararon los diagnósticos realizados con la entrevista PRIME-MD con el diagnóstico llevado a cabo por profesionales de salud mental. Sobre una muestra de 1000 personas, la concordancia fue moderada,  $\kappa = 0.71$  (Spitzer et al., 1994).

En otro estudio, se aplicaron las entrevistas PRIME-MD de manera telefónica o presencial, así como la entrevista SCID-I de manera telefónica. Se encontraron valores de  $\kappa = 0.67$  en la versión telefónica y  $\kappa = 0.70$  en la versión presencial utilizando como criterio el diagnóstico de la entrevista SCID-I (Kobak, Dottl, y Greist, 1997).

### **Adaptación española**

En la versión española del cuestionario PRIME-MD (Baca et al., 1999) se ha estudiado su validez diagnóstica utilizando la entrevista SCAN como instrumento de referencia. La concordancia para el conjunto de los trastornos fue  $\kappa = 0.45$  y para los trastornos del estado de ánimo,  $\kappa = 0.50$ . Con una sensibilidad del 72.2, una especificidad del 86.0 (Baca et al., 1999).

#### **- Protocolo de Entrevista Diagnóstica (DIS; *Diagnostic Interview Schedule*).**

Se trata de una entrevista totalmente estructurada, que puede ser aplicada tanto por personal clínico como no clínico (Epidemiology and Prevention Research Group, 2010). Está basado por completo en el DSM; de hecho, el DIS-II se desarrolló a la vez que el DSM-III y fue modificada cuando se publicó el DSM-III-R para adaptarse a sus cambios (Robins et al., 2001). En su última versión, DIS-IV (Robins et al., 1995), la entrevista valora el diagnóstico de los trastornos mentales del DSM-IV. Sus objetivos son ofrecer una historia vital de los síntomas del paciente para realizar el diagnóstico de los trastornos que padece, así como ofrecer una valoración de la sintomatología durante los últimos 12 meses y en la actualidad.

El DIS tiene una amplia cobertura, con 19 módulos que valoran más de 30 diagnósticos del DSM-IV, entre otros los trastornos depresivos. Aunque existen indicaciones para que el evaluador pueda saltarse las preguntas cuando se ha descartado la presencia del trastorno, se puede continuar la aplicación del resto de preguntas para valorar la sintomatología subumbral. Además, incluye información sobre el curso de algunos trastornos (como los trastornos depresivos), el deterioro social, laboral o de las relaciones sociales, los tratamientos aplicados, los factores de riesgo y características demográficas, los comportamientos saludables y los

indicadores sociales. Es una entrevista relativamente larga de administrar, con un tiempo aproximado de entre 90-120 minutos.

Existe una versión breve de esa entrevista (Quick DIS-III-R) que ha mostrado buenos datos de sensibilidad y especificidad (Bucholz, Marion, Shayka, Marcus y Robins, 1996).

#### **Fiabilidad**

Los datos de fiabilidad provienen en su mayoría de las primeras versiones del DIS, con índices kappa en los estudios test re-test aceptables, con kappas por encima de 0,5 (Robins et al., 1981 y Robins, Helzer, Ratcliff, Seyfried, 1982, en Summerfeldt et al., 2011).

El principal estudio de fiabilidad de la última versión, el DIS-IV, se realizó sobre una muestra de sujetos con diagnóstico de abuso de sustancias. En este caso el grado de acuerdo entre los diagnósticos realizados en dos momentos distintos fue aceptable, encontrándose una fiabilidad test-retest de 0.67 en el diagnóstico de los episodios de depresión mayor (Dascalu, Compton, Horton y Cottler, 2001; Horton, Compton y Cottler, 1998, en Epidemiology and Prevention Research Group, 2010).

#### **Validez**

Los datos sobre la validez del DIS son peores de lo esperado, de manera que se ha señalado que el DIS-IV no es adecuado como método principal de diagnóstico en la práctica clínica, puesto que se han encontrado que en general los resultados de los estudios sobre su validez “de algún modo decepcionantes” (Nezu, Nezu, Friedman y Lee, 2009, p. 45). Por ejemplo en un estudio se encontró, en el caso de los trastornos depresivos, una concordancia pobre de kappa = 0.20 y una

sensibilidad baja (0.29), tomando como criterio el diagnóstico realizado mediante la DIS-IV y la SCAN en una muestra de 349 sujetos (Eaton, Neufeld, Chen y Cai, 2000).

### **Adaptación española**

Existe una versión española del DIS-III-R, adaptada a partir de las versiones en castellano de la entrevista original que ha mostrado datos de fiabilidad y validez positivos (Bethencourt, Peñate, González y Fumero, 1996, Pablo Alcázar et al., 1992).

#### **- Protocolos para la Evaluación Clínica en Neuropsiquiatría (SCAN; *Schedules for Clinical Assessment in Neuropsychiatry*)**

Esta entrevista o, mejor dicho, conjunto de entrevistas, también conocido como Cuestionarios para la Evaluación Clínica en Neuropsiquiatría, se trata de un instrumento desarrollado por la OMS para población psiquiátrica. Los autores del SCAN señalan que su objetivo es “proporcionar los medios completos, precisos y técnicamente específicos para describir y clasificar los fenómenos clínicos a fin de poder compararlos. Hacer comparaciones es el centro de las actividades clínicas, educativas y científicas”. Por tanto la entrevista trata de “promover observaciones clínicas de alta calidad (...), mejorar los conceptos clínicos mediante la enseñanza del lenguaje clínico común (...), y acelerar la acumulación de conocimientos” (Wing, 1998, p. 25).

La versión actual del instrumento se denomina SCAN 2.1. Está compuesta por 4 elementos diferenciados. La parte principal es una entrevista semiestructurada denominada *Present State Examination* (Examen del Estado Actual), que actualmente está en su 10ª versión (PSE- 10) y que tiene una amplia cobertura de los principales trastornos mentales de la CIE-10 y el DSM-IV. Además,

del PSE, el SCAN dispone de un glosario detallado de definiciones para ser utilizado en la evaluación de la sintomatología detectada en la entrevista, un listado de preguntas para realizar a informantes sobre el estado del paciente y un historial clínico opcional para la recogida de información complementaria.

En el PSE los síntomas no están organizados en función del diagnóstico a asignar al paciente, sino por su similitud. Al aplicarlo, el evaluador elige el periodo de tiempo en el que quiere valorar la situación del sujeto, desde el estado precedente (un mes antes de la aplicación de la entrevista) a la historia de trastornos ocurridos hasta la fecha o en un periodo representativo de tiempo.

La aplicación del SCAN debe ser realizada por especialistas en salud mental entrenados para ese propósito, aunque se han descrito también buenas cualidades psicométricas del instrumento en su aplicación por personal no especializado entrenado (Brugha, Neinhuis, Bagchi, Smith y Metzler, 1999, en Summerfeldt et al., 2011). La duración del conjunto de entrevistas del SCAN es de aproximadamente 90 minutos. Existe una aplicación de ordenador (CATEGO-5) para realizar diagnósticos según las clasificaciones DSM-IV y CIE-10 que incluye una escala de gravedad de síntomas según las secciones a valorar. Además, se ha publicado una versión breve, mini-SCAN, con buenos resultados psicométricos preliminares (Nienhuis, Rijnders, de Jonge y Wiersma, 2010).

#### **Fiabilidad**

En la última versión, el SCAN 2.1, se ha encontrado, en el caso de los trastornos depresivos, una fiabilidad test-retest (con una semana de separación) moderada ( $\kappa = 0.52$ ) y una fiabilidad interjueces alta (porcentaje de acuerdo del 87%), con una sensibilidad del 59% y una especificidad del 90% (Rijnders et al., 2000).

### **Validez**

Se ha comentado anteriormente cómo la entrevista SCAN ha sido utilizada para la validación de la entrevista CIDI, con unos resultados moderados. Además, se han encontrado datos de validez en general aceptables con respecto a la entrevista CIS-R (Jordanova et al., 2004).

### **Adaptación española**

Existe una adaptación al español de la que son autores Vázquez-Barquero et al., (1994) y que ha obtenido índices de fiabilidad aceptables y similares a los obtenidos en otros países.

### **Conclusiones acerca de las entrevistas estructuradas y semiestructuradas**

La fiabilidad de las entrevistas diagnósticas de los trastornos depresivos es en general entre moderada y buena. Su contenido se ajusta por lo general a los criterios de los sistemas diagnósticos más validados y utilizados, esto es, el DSM-IV y la CIE-10, pero se necesita una mayor investigación sobre su validez. Son estudios complejos en su realización y en su interpretación, en los que hasta ahora se han encontrado datos de acuerdo moderados.

Por otro lado, no existe, en general, una investigación adecuada sobre las propiedades de fiabilidad y validez de la mayoría de las adaptaciones españolas de las entrevistas diagnósticas. Por ejemplo, la validación de la entrevista CIDI 3.0 se realizó con una versión adaptada de la entrevista SCID-I Versión de Investigación que no ha sido publicada y de la que, por tanto, se desconocen sus propiedades psicométricas. Tratándose de una adaptación de la entrevista cuya fiabilidad es considerada “menos que perfecta” (Segal et al., 2006), sería conveniente realizar estudios sobre población española. Tampoco se conocen en España las propiedades de fiabilidad y validez de la entrevista SCID-I-IV Versión Clínica (First



et al., 1999), a pesar de ser la entrevista de referencia para el diagnóstico de los trastornos del estado de ánimo. Otro ejemplo es la adaptación al castellano de la entrevista MINI, que se ha validado a través de los diagnósticos que los clínicos realizan mediante entrevistas no estructuradas o libres (Bobes, 1998).

### 3.3. Cuestionarios, escalas e inventarios para los trastornos depresivos

La utilización de cuestionarios, inventarios y escalas en el campo de los trastornos depresivos es útil para asignar pacientes a condiciones experimentales, como apoyo en la toma de decisiones sobre la necesidad de terapia, para la medición de la gravedad de la sintomatología y de la eficacia de los tratamientos, y como ayuda en la detección de los trastornos.

Los clínicos y los investigadores “tienen disponibles test de *screening* breves, flexibles y viables” para los trastornos depresivos, por lo que se ha recomendado su uso en EE. UU. especialmente para facilitar la detección y para cuantificar la gravedad de la sintomatología depresiva (U.S. Preventive Services Task Force, 2009).

Existe en atención primaria un gran debate sobre la utilización sistemática de instrumentos de *screening* o cribado para la detección y posterior tratamiento de los trastornos depresivos, con datos aún no definitivos sobre la eficacia de su introducción (Gilbody, Bower, Fletcher, Richards y Sutton, 2006; MacMillan, Patterson y Wathen 2005; National Institute for Clinical Excellence, 2004). Cuestionarios como el PHQ-9 (desarrollados y financiados por farmacéuticas como Pfizer) están siendo popularizados en países como EE. UU. e Inglaterra con el objetivo de convertirlos en un instrumento accesible a toda la población, que permita a los usuarios conocer de manera rápida si cumplen o no los criterios para la depresión mayor y monitorizar sus síntomas de manera dimensional. Existe una

gran polémica sobre estas recomendaciones en atención primaria, pero la utilidad de estos instrumentos en los servicios de salud mental parece evidente, puesto que proporcionan información útil más allá del diagnóstico.

En general, se han encontrado altas correlaciones entre los instrumentos de medida que aquí se describen. En parte, este hecho es fácil de explicar a partir de su validez de contenido, que gracias a la relativa estabilidad de los criterios diagnósticos del trastorno depresivo mayor durante las últimas décadas, ha servido de guía permanente para la elaboración de los diferentes ítems de los cuestionarios.

Los cuestionarios, inventarios y escalas que se van a revisar en este capítulo se han seleccionado por varias razones:

- **Por su frecuencia de uso.** Se han desarrollado más de 280 instrumentos de evaluación de la depresión durante los últimos 80 años, de entre los cuales solo unos 70 son objeto de publicaciones posteriores distintas a las que sirvieron para su elaboración y presentación (Santor, Gregus y Welch, 2006). Los tests de evaluación de la gravedad de los síntomas depresivos más utilizados durante los últimos años han sido desarrollados en su mayoría hace más de dos décadas: el BDI, la HDRS (o HAMD), la MADRS, el SCL-90 y el CES-D (Santor et al., 2006).
- Por la **disponibilidad de datos psicométricos adecuados** para su uso.
- Por ser los **instrumentos más citados en artículos publicados en revistas científicas**. Según una reciente revisión de Lipsman y Lozano (2011) que analizaba los artículos científicos sobre la depresión que habían sido más citados en las revistas científicas del área, se podría estimar que los cuestionarios, inventarios y escalas de depresión para adultos que más se

citan, dado que los estudios sobre los mismos son a su vez los más citados, son el Inventario de Depresión de Beck (BDI; *Beck Depression Inventory*), la Escala de Valoración de la Depresión de Hamilton (HAMD o HDRS; *Hamilton Rating Scale for Depression*), la Escala Hospitalaria de Ansiedad y Depresión (HADS; *Hospital Anxiety and Depression Scale*) y la Escala Autoaplicada para la Depresión de Zung (SDS o SRDS; Zung Self-Rating Depression Scale) (véase la Tabla 3.7).

**Tabla 3.7.** Estudios sobre cuestionarios, inventarios y escalas que han sido más citados en el campo de los trastornos depresivos (modificado de Lipsman y Lozano, 2011)

Estudio	Citas	Instrumento
Beck et al., 1961	15 590	BDI
Hamilton, 1960	12 047	HAMD o HDRS
Zigmond y Snaith, 1983	8 987	HADS
Zung, Richards y Short, 1965	4 532	SDS o SRDS
Beck, Steer, Garbin, 1988	4 423	BDI
Hamilton, 1967	4 262	HAMD
Montgomery, Asberg, 1979	3 759	MADRS
Lovibond y Lovibond, 1995	1 386	DASS
Bjelland, Dahl y Haug, 2002	1 162	HADS
Beck y Beamesderfer, 1974	1 119	BDI
Herrmann, 1997	1 013	HADS

*Nota.* BDI: Inventario de Depresión de Beck (*Beck Depression Inventory*); HAMD o HDRS: Escala de Valoración de la Depresión de Hamilton (*Hamilton Rating Scale for Depression*); HADS: Escala Hospitalaria de Ansiedad y Depresión (*Hospital Anxiety and Depression Scale*); SDS o SRDS: Escala Autoaplicada para la Depresión de Zung (Zung Self-Rating Depression Scale)

Los cuestionarios, escalas e inventarios que se van a revisar a continuación se pueden agrupar en dos grandes categorías: aquellos que han sido diseñados específicamente para realizar el diagnóstico de los trastornos depresivos y aquellos otros que no han sido diseñado específicamente para tal fin, sino que normalmente han sido diseñados para evaluar la gravedad de la sintomatología depresiva,

aunque también se utilizan para identificar posibles casos de pacientes con trastornos depresivos, es decir, para labores de cribado (*screening*), o incluso han sido diseñados para estas últimas labores, pero, en todo caso, nunca se han pensado para realizar diagnósticos.

### 3.3.1. Cuestionarios, escalas e inventarios diagnósticos

- **Cuestionario de Salud del Paciente-9 (PHQ-9; *Patient Health Questionnaire-9*)**

El cuestionario PHQ forma parte de la entrevista PRIME-MD (Spitzer et al., 1994). Su objetivo inicial es el de realizar diagnósticos sobre los trastornos mentales de manera fiable y válida en pacientes de atención primaria en un tiempo breve (Spitzer, Kroenke y Williams, 1999). El módulo de depresión de este cuestionario se denomina PHQ-9 (Kroenke, Spitzer, y Williams, 2001), y en él el sujeto debe valorar cada uno de los 9 síntomas principales del episodio depresivo mayor mediante cuatro alternativas de respuesta: “en absoluto”, “en algunos días”, “más de la mitad de los días” o “casi todos los días”, centrándose en cómo se ha sentido durante las dos últimas semanas. El objetivo es conseguir una valoración dimensional de la gravedad de la depresión del paciente, a la vez que siguiendo los algoritmos de decisión correspondientes se asigna el diagnóstico. Además, el sujeto debe valorar hasta qué punto estas dificultades le han generado interferencia con el trabajo, la actividad diaria o las relaciones sociales.

El experto que corrige el cuestionario debe tener en cuenta las posibles explicaciones médicas a la sintomatología, su funcionamiento normal y los episodios maniacos previos.

Esta escala ha sido propuesta por la comisión de elaboración de los trastornos depresivos del DSM-5 como instrumento para la medición de la gravedad

del trastorno. Las categorías de gravedad de la depresión y sus correspondientes puntos de corte que se han previsto son: depresión mínima (1-4), depresión leve (5-9), depresión moderada (10-14), depresión moderadamente grave (15-19) y depresión grave (20-27) (APA, 2011).

Para la validación del PHQ-9, se aplicó por parte de profesionales especializados una entrevista diagnóstica independiente a 6000 personas de una población clínica. Los datos señalaron en el caso del trastorno depresivo mayor una buena sensibilidad (0.88) y especificidad (0.88). Sin embargo se observó un bajo poder predictivo positivo (entre 0.31 y 0.51 dependiendo del punto de corte seleccionado) (Kroenke et al., 2001).

Existe una versión breve del PHQ-9, el PHQ-2, que tan solo incluye dos preguntas relacionadas con la tristeza y la anhedonia que también se valoran de 0 a 3. El punto de corte que sugieren los autores para diagnóstico de un trastorno depresivo es de 3, para el cual se ha encontrado buenas propiedades diagnósticas (sensibilidad de 0.83 y especificidad de 0.90) (Kroenke, Spitzer y Williams, 2003).

Gilbody, Richards, Brealey y Hewitt (2007) llevaron a cabo un estudio metaanalítico en el año 2007 en el que se incluyeron 17 publicaciones sobre las propiedades diagnósticas del PHQ-9 y del PHQ-2. Estos investigadores encontraron para el PHQ-9 una sensibilidad del 0.80 y una especificidad de 0.92 con un punto de corte de 10 puntos, sin una variación significativa cuando se utilizaba el algoritmo de decisión para realizar el diagnóstico según las respuestas dadas al PHQ-9. En los estudios revisados se proponían distintos puntos de corte debido a su mayor eficacia diagnóstica, variando desde los 9 puntos a los 13. Con respecto al PHQ-2, los autores señalaban que no habían encontrado suficientes investigaciones acerca de sus propiedades diagnósticas como para sacar

conclusiones firmes, aunque en principio tendría una buena sensibilidad y un bajo valor predictivo positivo.

Patten y Schopflache (2009, p. 32) señalan que:

“El rol apropiado de los instrumentos como el PHQ-9 no es el de diagnosticar la depresión o evaluar la necesidad de tratamiento, sino identificar personas que necesitan una evaluación más profunda. Los resultados de nuestra publicación refuerzan esa idea, debido a que los síndromes depresivos identificados por el PHQ-9 parecen representar una grupo heterogéneo”.

Además, estos autores afirman que el PHQ-9 no es tan útil para proporcionar información sobre el diagnóstico de manera categorial, puesto que aunque las puntuaciones muy bajas o muy altas sí tienen un gran poder predictivo, las puntuaciones intermedias presentan más dudas en su interpretación, independientemente de si cumplen o no con la sintomatología marcada por los algoritmos de decisión. Esta utilización del PHQ-9 como una medida dimensional más que categorial está siendo cada vez más aceptada, a la vez que otros autores han encontrado en este sentido buenos datos de validez (Kroenke et al, 2001).

En el año 2010 se llevó a cabo un estudio en Nueva Zelanda sobre una muestra de 2642 personas, a las que se les aplicó el PHQ-9 y, como criterio diagnóstico, la entrevista CIDI (Arroll et al., 2010). En dicho estudio, se encontraron buenos datos de sensibilidad (0.74) y especificidad (0.92) para el punto de corte 10 (Arroll et al., 2010). En este sentido los autores señalan que este punto de corte es adecuado, por lo que “una reevaluación de los criterios de corrección originales para la depresión mayor es necesaria, puesto que la simple suma de las puntuaciones identifica más pacientes que el método original” (Arroll et al., 2010, p. 353). Por otro

lado, en ese estudio el PHQ-2 mostró buena sensibilidad, pero baja especificidad (Arroll et al., 2010).

#### **Adaptación española**

La adaptación española del PHQ-9 se llevó a cabo en una muestra de 1003 pacientes hospitalarios (Diez-Quevedo, Rangil, Sanchez-Planell, Kroenke y Spitzer, 2001). Se compararon las puntuaciones de los sujetos en este instrumento con respecto a las puntuaciones del BDI, medidas de estatus funcional y utilización de asistencia clínica. Se obtuvieron datos que señalaban una buena validez del instrumento, con un acuerdo diagnóstico con respecto a un experto profesional de  $\kappa = 0.75$ , sensibilidad de 87% y especificidad del 88%. Además, se encontró una correlación significativa del instrumento con el BDI-II (Diez-Quevedo et al., 2001).

#### **- Inventario Diagnóstico para la Depresión (DID; *Diagnostic Inventory for Depression*)**

El DID es un inventario autoaplicado diseñada específicamente para “el diagnóstico provisional del trastorno depresivo mayor según los criterios DSM-IV” (Zimmerman, Sheeran y Young, 2004, p. 88). El instrumento valora los síntomas incluidos en el DSM-IV para el periodo temporal de las dos últimas semanas, así como el nivel de deterioro funcional y la calidad de vida del paciente. Los ítems deben ser valorados por el paciente de 0 a 4 (desde *sin dificultad* hasta *dificultad extrema*), siendo una puntuación mayor de 2 considerada como presencia del síntoma.

En un estudio realizado por los autores del inventario, se aplicó a 626 personas un protocolo de evaluación que incluía, además del DID, la entrevista SCID (junto con algunas cuestiones añadidas de la entrevista SAD) y la escala de

heteroevaluación HRSD (Zimmerman et al., 2004). Los evaluadores, que habían recibido entrenamiento previo en la aplicación de la entrevista diagnóstica, realizaron el diagnóstico sin conocer las respuestas de los pacientes al DID. Los datos señalaron buena consistencia interna (alfa de Cronbach = 0.91) del DID. La fiabilidad test-retest del inventario ( $N = 101$ ) fue superior a 0.75, teniendo en cuenta que una parte de los participantes contestó las dos veces al DID en un mismo día (Zimmerman et al., 2004).

Es importante destacar que este inventario realiza los diagnósticos de depresión mayor a través de varios sistemas de corrección. En concreto, se han analizado cuatro modelos distintos para llegar a un diagnóstico de depresión mayor:

1. Mediante la aplicación de los algoritmos de decisión del DSM-IV a los síntomas para determinar si éstos se cumplen.
2. Teniendo en cuenta el cumplimiento de los síntomas, junto con la duración de los mismos.
3. Añadiendo a los síntomas el grado de interferencia que provocan en el paciente.
4. Valorando tanto los síntomas como su duración y el grado de interferencia que producen.

En todos los casos se han encontrado buenas propiedades de eficacia diagnóstica (véase la Tabla 3.8.), con valores de sensibilidad, poder predictivo positivo y poder predictivo negativo por encima de 0.75 y kappas entre 0.61 y 0.67 (Zimmerman et al., 2004).



**Tabla 3.8.** Propiedades diagnósticas del DID en función del sistema de corrección seleccionado (Zimmerman et al., 2004)

<b>Sistema de corrección</b>	<b>Nº de casos identificados</b>	<b>Sens.</b>	<b>Espec.</b>	<b>PPP</b>	<b>PPN</b>	<b>Kappa</b>
Algoritmos DSM-IV	360	90.1	77.0	80.6	88.0	0.67
Cumplimiento de síntomas + duración de síntomas	311	80.1	82.6	83.0	79.7	0.63
Cumplimiento de síntomas + grado de interferencia	337	79.9	79.9	81.9	84.1	0.66
Cumplimiento de síntomas + duración de síntomas + grado de interferencia	294	84.5	84.5	84.0	77.4	0.61

*Nota.* Sens.: sensibilidad; Espec.: especificidad; PPP: poder predictivo positivo; PPN: poder predictivo negativo.

Los autores de ese estudio destacan que a pesar de estas buenas cualidades, el instrumento no debe ser utilizado como diagnóstico definitivo, puesto que no puede distinguir entre el trastorno depresivo mayor y una reacción de duelo, una depresión debida al consumo de medicamentos, una enfermedad médica, el abuso de sustancias o la depresión que sucede en el curso de un trastorno psicótico (Zimmerman et al., 2004).

El cambio de estrategia que estos autores proponen al utilizar los algoritmos de decisión de los manuales diagnósticos, trata de solucionar un problema frecuente en la investigación: los estudios utilizan distintos puntos de corte para la detección de los trastornos que además no están sometidos a una toma de decisiones racional. Pero, en este sentido, ahora el problema se ha convertido en qué algoritmo de decisión es el más adecuado (Sheeran y Zimmerman, 2002).

Sheeran y Zimmerman (2002) analizaron, en una muestra de 788 participantes, la eficacia diagnóstica de dos sistemas de corrección del DID, uno basado en un punto de corte teniendo en cuenta la sintomatología y otro basado en el algoritmo de decisión del DSM-IV. Para ello, realizaron de nuevo una comparación de este procedimientos de corrección del DID con el diagnóstico realizado a través de la SCID, sin que los terapeutas conocieran las puntuaciones del DID (véase la Tabla 3.9.). El análisis de la curva ROC reveló buenas propiedades diagnóstica para los dos sistemas de corrección del instrumento, sin que existieran diferencias significativas entre ellos (área de la curva ROC = .90 en el caso del DID corregido mediante un punto de corte basado en la sintomatología y área de la curva ROC = 0.89 en el caso del DID corregido sobre la base del algoritmo de decisión del DSM-IV).

**Tabla 3.9.** Estadísticos de eficacia diagnóstica del DID calculados a partir de dos sistemas de corrección: mediante la suma de la sintomatología o mediante el algoritmo de decisión del DSM-IV (Sheeran y Zimmerman, 2002)

<b>Estadísticos</b>	<b>Punto de corte mediante sintomatología</b>	<b>Algoritmos de decisión DSM-IV</b>
Sensibilidad	0.91	0.85
Especificidad	0.70	0.73
PPP	0.76	0.76
PPN	0.88	0.82
Kappa	0.62	0.58

*Nota.* PPP: poder predictivo positivo; PPN: poder predictivo negativo.

En conclusión, los datos de eficacia diagnóstica recogidos en la Tabla 3.9 no apoyan la superioridad de los algoritmos de decisión sobre la suma de los síntomas a la hora de realizar un diagnóstico de trastorno depresivo mediante la aplicación de cuestionarios, escalas o inventarios autoaplicados, al menos mediante el DID. De hecho, en el estudio de Sheeran y Zimmerman (2002) el único indicador en el que

aparecieron diferencias significativas fue en el estadístico kappa y a favor del sistema de punto de corte tradicional (sintomatología).

### **Adaptación española**

No ha sido posible localizar ninguna traducción validada en castellano del DID.

### **3.3.2. Cuestionarios, escalas e inventarios de cribado**

- **Escala de Hamilton para la Depresión (HAMD, HRSD o HDRS; *Hamilton Rating Scale for Depression o Hamilton Depression Rating Scale*).**

La HAMD, HRSD o HDRs (Hamilton, 1960) ha sido durante años el instrumento de evaluación clínica más utilizado y la referencia más importante en la medición de la gravedad de la depresión. Originalmente fue diseñada para medir la sintomatología depresiva de los pacientes diagnosticados con depresión clínica y se ha usado con mucha frecuencia en investigación, especialmente como método de medida de la eficacia de los psicofármacos.

La HAMD es un instrumento de heteroevaluación en el que el clínico debe valorar la gravedad de la depresión del paciente después de llevar a cabo una entrevista con el mismo. Hamilton (1960) aconsejaba que la duración de dicha entrevista fuera de al menos 30 minutos y que estuviera centrada en la situación del paciente durante la última semana. Existen varias versiones del instrumento que difieren en el número de ítems que la componen: 17, 21, 24 y 27 ítems, y, desafortunadamente, los investigadores no siempre informan de qué versión han utilizado en sus estudios. De la versión de 21 ítems, en 17 de ellos los clínicos deben valorar la frecuencia e intensidad de los síntomas con una puntuación de 0 a 4 (9 ítems) o con una puntuación de 0 a 2 (8 ítems). Se han realizado varias

propuestas para los puntos de corte del cuestionario que permiten distinguir niveles o categorías de gravedad de la depresión. Las categorías de gravedad y los puntos de corte propuestos por Bech (1996) son los siguientes: no depresión (0-7), depresión menor (8-12), menos que depresión mayor (13-17), depresión mayor (18-29) y más que depresión mayor (30-52).

Este instrumento mide fundamentalmente síntomas comportamentales y somáticos de la depresión como el insomnio, el retardo psicomotor, la agitación, la ansiedad, la pérdida de apetito, la pérdida de peso o dolor muscular. En versiones posteriores, se añadieron síntomas cognitivos como desesperanza o sentimiento de inutilidad. Bagby, Ryder, Schuller y Marshall (2004) destacan que algunos de los síntomas que mide la HAMD no se corresponden por completo con los criterios DSM-IV, por lo que algunos autores han puesto en duda su validez de contenido.

#### **Fiabilidad**

Un metaanálisis sobre los estudios de fiabilidad de la HAMD durante los últimos 49 años transcurridos desde su aparición, ha mostrado buenos datos de consistencia interna, fiabilidad inter-jueces y fiabilidad test-retest (Trajković, et al., 2011).

#### **Validez**

Respecto a la validez diagnóstica, los puntos de corte más eficaces determinados por los investigadores a través del análisis de la curva ROC, varían entre 12 y 16 (Bagby, et al., 2004). Los estudios analizados por estos autores señalan buena sensibilidad, especificidad y poder predictivo negativo. En cambio, señalan resultados más variables y, en muchos casos pobres, en cuanto al poder predictivo negativo. La validez convergente de este instrumento es buena, habiéndose encontrado correlaciones con otros cuestionarios que miden depresión

como, por ejemplo, el BDI que varían según los estudios entre 0.48 a 0.89 (Bagby et al., 2004). Sin embargo, como se ha comentado antes, la validez de contenido de la HAMD con respecto a los criterios del DSM-IV ha de ponerse en duda (Bagby et al., 2004). La multidimensionalidad del HAMD, así como su falta de correspondencia con los criterios DSM-IV, hacen que estos autores reclamen que “es tiempo de retirar la escala de depresión de Hamilton. Se necesita dar un paso adelante y seleccionar un nuevo patrón oro que incorpore modernos métodos psicométricos y definiciones contemporáneas de la depresión” (Bagby, et al, 2004, p. 2163). Zimmerman, Posternak y Chelminski (2005) también han señalado la necesidad de abandonar el uso del HAMD como el “patrón oro” para la medida de la depresión.

#### **Adaptación española**

La versión de la HAMD de 17 ítems se validó en una muestra pequeña de pacientes hospitalizados (Ramos-Brieva y Cordero, 1986; Ramos-Brieva, Cordero y Yáñez, 1994), con buenos datos de consistencia interna y fiabilidad inter-jueces. Estos datos fueron confirmados en un estudio posterior con una muestra más amplia ( $N = 168$ ) para la adaptación de las escalas de 21, 17 y 6 ítems (Bobes et al., 2003). Se encontraron buenos datos de consistencia interna y validez para discriminar entre distintos niveles de gravedad de los síntomas, incluso para la versión breve de 6 ítems, así como buena capacidad para medir el cambio tras el tratamiento farmacológico. Todos los sujetos de la muestra estaban diagnosticados de trastorno depresivo o trastorno adaptativo con ánimo deprimido, por lo que no se valoró su validez discriminante ni sus características como instrumento diagnóstico (Bobes et al., 2003). Otros autores también han señalado que la HAMD es un instrumento válido para la valoración del cambio terapéutico en su versión española (Ballesteros et al., 2007).

**- Inventario de Depresión de Beck-II (BDI-II; *Beck Depression Inventory-II*)**

***II)***

El BDI-II es un instrumento diseñado para la evaluación de la sintomatología depresiva en adultos y adolescentes (con una edad mínima de 13 años). Es un instrumento especialmente interesante para la medición del cambio durante el proceso terapéutico, uno de los más conocidos en la práctica clínica (Feliciano y Gum, 2010) y el más citado en la investigación sobre la depresión (Santor et al., 2006).

Se trata de un instrumento de autoinforme autoaplicado que está compuesto por 21 ítems. En cada uno de ellos el sujeto debe seleccionar, entre un conjunto de cuatro alternativas ordenadas de menor (0) a mayor gravedad (3) (salvo dos ítems que tienen cada uno 7 alternativas, pero que también se puntúan de 0 a 3), la frase que mejor describe su estado durante las últimas dos semanas. Aunque el BDI-II es un instrumento de papel y lápiz, también se puede aplicar de manera oral si el paciente tiene dificultades visuales o de concentración. El tiempo de administración del BDI-II es de 5-10 minutos (Beck, Steer y Brown, 1996).

El BDI-II es la segunda edición del BDI original publicado en 1961 (Beck et al., 1961), y tanto en esta versión original o primera edición (también conocida como BDI-I) como en su posterior modificación de 1978 o BDI-IA (Beck et al., 1979; Beck y Steer, 1987, 1993) como en su segunda edición, el instrumento recoge un gran porcentaje de los síntomas descritos por las distintas versiones del DSM para la depresión mayor. Sin embargo, el solapamiento entre los síntomas que miden el BDI o el BDI-IA y los criterios sintomáticos del DSM-IV no es completo, por lo que para la elaboración del BDI-II (Beck et al., 1996) se realizó una profunda modificación en la redacción del contenido de los ítems y se añadieron ítems nuevos para acercar el contenido del inventario a los criterios diagnósticos del

trastorno depresivo mayor del DSM-IV (APA, 2002). Así, por ejemplo, en el BDI-II se incluyeron ítems para la valoración de síntomas como la agitación psicomotora o las dificultades de concentración.

Beck et al. (1996) señalan que el BDI-II no debe ser utilizado como instrumento diagnóstico. Consideran que los cuestionarios de autoinforme están demasiado influidos por los sesgos de respuesta, lo que provoca que en ocasiones los pacientes respondan por factores distintos a la descripción de sus propios síntomas, por lo que “La determinación de la depresión y el establecimiento de un diagnóstico requiere el examen realizado por un clínico” (Beck, Steer y Brown, 2011a, p. 24). El BDI-II no contiene los criterios de exclusión fijados por el DSM-IV y no tiene en cuenta en su corrección, por ejemplo a través de algún algoritmo específico, que deben cumplirse al menos uno de los dos criterios sintomáticos de tristeza o anhedonia para poder realizar el diagnóstico del trastorno depresivo mayor. A pesar de ello, son numerosos los estudios que han establecido puntos de corte para la detección de los trastornos depresivos (Zimmerman et al., 2004).

En el manual del BDI-II (Beck et al., 1996), se describen los pasos dados para la elaboración de los puntos de corte que permiten distinguir distintos niveles de gravedad de la sintomatología depresiva. En el estudio los pacientes de la muestra fueron diagnosticados por clínicos con la ayuda de la entrevista SCID-I, y los pacientes diagnosticados con un trastorno depresivo mayor actual fueron a su vez clasificados según la gravedad de su depresión en depresión leve, moderada o grave. A partir de estos juicios diagnósticos, se realizaron análisis de las curvas ROC para las categorías de no deprimidos, depresión leve, depresión moderada y depresión grave, y se determinaron los rangos que aparecen en la Tabla 3.10.

**Tabla 3.10.** Rangos recomendados para la interpretación del BDI-II (Beck et al., 1996)

Puntuación total BDI-II	Rango
0 – 13	Mínima
14 – 19	Leve
20 – 28	Moderada
29 – 64	Grave

Beck et al. (1996) señalan que cada usuario debe utilizar los puntos de corte más adecuados para su muestra y sus propósitos. En el caso de los puntos de corte establecidos en su estudio, se ha hecho más hincapié en la sensibilidad que en la especificidad para dotarles de unas buenas cualidades como instrumento de cribado o *screening*.

Se ha señalado una estructura de dos factores de este instrumento: un factor de sintomatología somática-emocional y otro de sintomatología cognitiva (Beck et al., 1996), y en la adaptación española se ha encontrado soporte para esta estructura (cognitivo-afectiva y somático-motivacional; Sanz y García Vera, 2009).

### Fiabilidad

En el estudio original de Beck et al. (1996), la consistencia interna fue alta (alfa de cronbach = 0,93 para estudiantes universitarios y 0,93 para pacientes ambulatorios).

En un análisis metaanalítico de las publicaciones sobre la consistencia interna del BDI (Yin y Fan, 2000) se incluyeron las versiones del BDI de 21 ítems. Dado que el estudio recogía publicaciones realizadas hasta el 1998, solamente se



incluyeron 2 estudios realizados con el BDI-II. Se encontraron buenos datos psicométricos con un alfa de Cronbach de 0,84. Además, encontraron índices de fiabilidad test-retest altas, en concreto se encontró un promedio de coeficientes de correlación test-retest = 0.72.

Dozois, Dobson y Ahnberg (1998) estudiaron la consistencia interna del BDI-II en una muestra de 1 022 estudiantes universitarios y encontraron un alfa de cronbach = 0,91.

#### **Validez**

La validez convergente del BDI-II es alta con la HAMD, habiéndose encontrado una correlación de 0.71 (Beck et al., 1996), y con la subescala de depresión de la *Depression Anxiety Stress Scale*, habiéndose informado en este caso de una correlación de 0.76 (Gloster, et al., 2008). Así mismo, la correlación del BDI-II es alta con la escala depresiva (0.89) y la escala de ansiedad (0.71) del SCL-90-R (Steer, Ball, Ranieri y Beck, 1997). Una crítica al BDI-II que se puede generalizar al conjunto de los cuestionarios, escalas e inventarios de depresión es que tienen dificultades para discriminar correctamente entre ansiedad y depresión, quizá precisamente por la alta comorbilidad existente entre estos trastornos.

En un reciente estudio (Titov et al., 2011), se compararon los instrumentos BDI-II y PHQ-9 en dos ensayos aleatorios controlados, en los que se obtuvo en ambos instrumentos una buena consistencia interna (para el PHQ-9, alfa = .74 y .81, y para el BDI-II, alfa = .87 y .90, en el pretratamiento y en el postratamiento, respectivamente). Además, se mostraron sensibles al cambio terapéutico y se encontraron altas correlaciones entre ellos. Sin embargo, ambos instrumentos difirieron en la clasificación de la gravedad de la sintomatología depresiva, puesto que el BDI-II clasificó un porcentaje mayor de sujetos como depresión grave.

Además, en este estudio se confirmó la estructura bifactorial tanto del BDI-II como del PHQ-9 (Titov et al., 2011).

Joiner et al, (2005.) han señalado que:

“El BDI-II es breve y sus propiedades métricas son bien conocidas, por lo que consideramos que es el instrumento de elección para la evaluación de la gravedad de la sintomatología depresiva y la valoración del cambio en la gravedad en muestras de pacientes no ambulatorios (por ejemplo en la valoración de la respuesta al tratamiento). Fue diseñado para este propósito y no fue diseñado específicamente para el establecimiento formal del diagnóstico de la depresión” (p. 274).

#### **Adaptación española**

Las propiedades psicométricas de la adaptación española del BDI-II se han estudiado en muestras de estudiantes, población general y pacientes clínicos con trastornos mentales (Sanz y García-Vera, 2013; Sanz, García-Vera, Espinosa, Fortún y Vázquez, 2005; Sanz, Navarro y Vázquez, 2003; Sanz, Perdigón y Vázquez, 2003), arrojando buenos datos de consistencia interna (coeficientes alfa entre 0.91 y 0.87), y recientemente dicha adaptación ha sido publicada por *Pearson Talent and Assessment España* (Beck et al., 2011a, adaptación de J. Sanz y C. Vázquez). Se han encontrado buenos datos de validez convergente respecto a otros instrumentos de evaluación de la depresión y aceptables datos sobre la validez discriminante, puesto que la correlación con instrumentos de medida general de la ansiedad es menor (Sanz, et al., 2005).

Sanz, Navarro y Vázquez (2003) recomiendan su uso como instrumento de medición de la gravedad de la sintomatología depresiva, así como su utilización como instrumento de cribado o *screening* para los trastornos depresivos. Estos

autores han realizado un análisis de la eficacia diagnóstica del BDI-II con respecto al diagnóstico de trastorno depresivo mayor, evaluado mediante la versión computerizada de la entrevista diagnóstica Quick DIS-III-R (Bucholz et al., 1996). Se encontraron buenos datos tanto de sensibilidad (92.9%) como especificidad (84.1%), con un índice de eficacia diagnóstica del 84.9% y un coeficiente kappa moderado (0.44). Estos datos fueron obtenidos con un punto de corte de 12, seleccionado mediante el análisis de la curva ROC (Sanz et al., 2003).

Existe también una versión breve del BDI-II denominada BDI-PC (*Beck Depression Inventory-Primary Care*) y, más recientemente BDI-FS (*Beck Depression Inventory-FastScreen*) y que fue diseñada para ser utilizada en atención primaria y, en general, en pacientes médicos (Steer, Cavalieri, Leonard y Beck, 1999; Beck, Steer y Brown, 2000). Para ello, se eliminaron del BDI-II los ítems que evaluaban síntomas depresivos que pudieran confundirse con síntomas físicos debidos a enfermedades médicas o a problemas de abuso de alcohol o de sustancias, y, por tanto, el BDI-PC o BDI-FS consta de tan sólo 7 ítems. Recientemente, también se ha publicado la adaptación española del BDI-FS (Beck, Steer y Brown, 2011b, adaptación española de J. Sanz, A. Izquierdo y M. P. García Vera), la cual presenta, en muestras de pacientes españoles de atención primaria, buenos índices psicométricos de fiabilidad de consistencia interna, de fiabilidad test-retest, de validez convergente, de validez discriminante y de validez factorial. Es más, esta adaptación española también muestra buenos índices de eficacia diagnóstica para detectar pacientes de atención primaria con trastornos depresivos cuando se utiliza como criterio el diagnóstico proporcionado por la SCID-I (Beck et al., 2011b; Izquierdo Elizo, 2011). En ese mismo sentido, apuntan los datos obtenidos con otra traducción española del BDI-FS realizada por Orive et al. (2010),

con la cual esos autores encontraron buenos datos de eficacia diagnóstica con respecto a una entrevista estructurada aplicada por clínicos.

- **Escala de Depresión del Centro de Estudios Epidemiológicos (CES-D; *Center for Epidemiologic Studies Depression Scale*)**

La CES-D es un instrumento breve (consta de 20 ítems), que tiene por objetivo la detección de personas con niveles altos de sintomatología depresiva en la población general (Radloff, 1997). Los sujetos deben contestar a las cuestiones puntuando de 0 a 3 con qué frecuencia han sentido durante la última semana los síntomas referidos en cada ítem, siendo 0 ausencia del síntoma y 3 casi todos los días. Su tiempo de aplicación es de aproximadamente 5 minutos.

Al contrario que el resto de cuestionarios, la CES-D hace mayor hincapié en la valoración del estado de ánimo del paciente. Incluye dos ítems sobre sintomatología vegetativa que valoran los problemas de apetito y sueño, pero no tiene ítems sobre ideación suicida, pérdida de energía o sintomatología inversa (p. ej., aumento de apetito o sueño). Por otra parte, incorpora elementos que no están habitualmente recogidos por los cuestionarios específicos para la valoración de los trastornos depresivos, como el miedo o la sensación de que las personas no son amables. Por tanto, su validez de contenido respecto del DSM-IV es menor que la de otros cuestionarios o inventarios nombrados con anterioridad.

### **Adaptación española**

Existen varias adaptaciones españolas de la CES-D. Una de ellas ha sido realizada por Soler, Pérez Sola, Puigdemot, Pérez-Blanco y Figueres (1997) con una muestra de 99 pacientes con trastornos del estado de ánimo, en la que encontraron buenos índices de fiabilidad de consistencia interna ( $\alpha = 0.99$ ) y de validez convergente con el BDI y la HAMD.

Otra adaptación española de la CES-D se ha llevado a cabo con una muestra española de personas mayores de 65 años que acudía a un centro de atención primaria (Zunzunegui et al., 1998). Los resultados con esta adaptación revelaban buenos índices de fiabilidad de consistencia interna ( $\alpha = 0.89$ ) y aceptables índices de eficacia diagnóstica para detectar pacientes con trastorno depresivo mayor o trastorno distímico en relación con el diagnóstico clínico de depresión (DSM-III-R) realizado por un psiquiatra de forma ciega (Zunzunegui et al., 1998).

Una tercera adaptación española ha sido examinada con una muestra de estudiantes universitarios (Vázquez, Blanco y López, 2007), encontrándose también una fiabilidad de consistencia interna satisfactoria ( $\alpha = 0.89$ ). Además, en este estudio, al comparar la eficacia diagnóstica del punto de corte de 26 en la CES-D con el diagnóstico realizado a través de la entrevista SCID-I-CV por un clínico ciego a la puntuación en la escala, se encontró una buena sensibilidad y especificidad (0.90 y 0.91, respectivamente) de la CES-D para detectar estudiantes con un diagnóstico de trastorno depresivo mayor (Vázquez et al., 2007).

Por último, una cuarta adaptación de la CES-D ha sido realizada por Gracia y Herrero (2004a, 2004b; véase también Herrero y Gracia, 2007) en muestras de la población general española, informándose de buenos índices de fiabilidad de consistencia interna ( $\alpha = 0.88$ ; Gracia y Herrero, 2004a). Es más, a partir de esta adaptación, Herrero y Gracia (2007) han adaptado también la versión de la CES-D de tan sólo siete ítems (CESD-7) que Santor y Coyne (1997) proponían en su búsqueda de una versión breve de la escala. Esta versión breve también ha mostrado en muestras de la población general española unos buenos índices de fiabilidad de consistencia interna ( $\alpha = 0.82$ ) y de validez factorial (Herrero y Gracia, 2007).

- **Escala de Autoevaluación de la Depresión de Zung (SDS; Self-Rating Depression Scale)**

Se trata de un instrumento autoaplicado y breve, desarrollado para la medición de la gravedad de la depresión (Zung, 1965). Consta de 20 ítems sobre los que el paciente debe indicar la frecuencia con la que experimenta cada uno de los síntomas explorados y que son valorados de 1 a 4, por lo que la puntuación total varía desde 20 a 80. Existe también un sistema de puntuación porcentual para la SDS basado en la transformación de las puntuaciones brutas en porcentajes sobre la puntuación bruta máxima de la escala. La SDS no se ha modificado desde su publicación, antes de la aparición del DSM-III, por lo que incluye dentro de su listado de ítems algunos que miden síntomas que actualmente no están incluidos dentro de los criterios diagnósticos para el trastorno depresivo mayor (p. ej., “Mi corazón se acelera más de lo habitual”), aunque, en general, tiene una buena validez de contenido (cubre síntomas como tristeza, anhedonia, dificultades para dormir, disminución del apetito, desesperanza y dificultad en la toma de decisiones).

La mitad de los ítems de la SDS están formulados de manera positiva y la otra mitad de manera inversa para evitar posibles sesgos de respuesta, y las instrucciones del instrumento piden al paciente que valore su estado en la actualidad (aunque en la adaptación española las instrucciones preguntan por la situación del paciente durante la última semana).

#### **Adaptación española**

Las propiedades psicométricas de la adaptación española del SDS, en cuanto a su fiabilidad de consistencia interna, validez convergente o validez factorial, son adecuadas y parecidas a las de la versión original tanto en muestras de pacientes psiquiátricos como de personas de la población general o de pacientes de atención primaria (Aragonès Benaiges, Masdéu Montalà, Cando

Guasch y Coll Borràs, 2001; Conde, Escribá e Izquierdo, 1970; Conde y Esteban, 1975; Izquierdo Elizo, 2011; véase también Conde y Franch, 1984).

En cuanto a la validación diagnóstica de esa adaptación española, Aragonès Benaiges et al. (2001) realizaron un estudio en una muestra de pacientes de atención primaria utilizando como criterio del diagnóstico de depresión mayor y distimia el módulo de trastornos del estado de ánimo de la SCID-I. Los resultados de este estudio fueron muy positivos, encontrándose buenos índices de sensibilidad (0.95), especificidad (0.74) y área bajo la curva ROC (0.93) (Aragonés Benaiges et al., 2001).

- **Escala de Valoración de la Depresión de Montgomery-Åsberg (MADRS; *MontgomeryÅsberg Depression Rating Scale*)**

La MADRS es una escala de heteroevaluación que debe ser aplicada por clínicos experimentados, aunque existen versiones autoaplicadas (Montgomery y Åsberg, 1979). Está compuesta por 10 ítems que evalúan los síntomas y la gravedad de la depresión, durante la última semana o los últimos 3 días. El contenido de los ítems del MADRS incluye la tristeza (tanto observada por el clínico como referida por el propio paciente), síntomas de tensión (incluida la ansiedad, el pánico o la angustia), disminución de la calidad del sueño y del apetito, dificultades de concentración, nivel de actividad, anhedonia, pensamientos pesimistas e ideación suicida. Como se puede observar el contenido de cuestionario es anterior al desarrollo del DSM-III y no ha sido actualizado en función de las modificaciones de las últimas versiones del DSM-IV o de la CIE-10.

Utilizando información de distintas fuentes, los clínicos puntúan cada ítem de 0 a 6. La puntuación total se calcula sumando las puntuaciones de cada ítem. Se han

recomendado distintos puntos de corte para su interpretación (Bobes, Portilla, Bascarán, Saiz y Bousoño, 2004).

### **Adaptación española**

La validación de la versión en español de la MADRS fue realizada en 2002 (Lobo et al., 2002), pero no se han estudiado sus cualidades como instrumento diagnóstico.

#### **- Cuestionario de los 90 Síntomas Revisado (SCL-90-R; *Symptom Checklist-90-Revised*)**

El SCL-90-R (Derogatis, 1977) es un cuestionario de autoinforme desarrollado para la evaluación de la psicopatología general. Aunque no es un instrumento inicialmente ideado para el diagnóstico, puede ser utilizado como cribado o *screening* en la detección de trastornos psicopatológicos. Este instrumento se centra en la medición de la gravedad de los síntomas ocurridos durante los últimos 7 días y en cada ítem se valora la presencia del síntoma de 0 a 4 (desde nada a extremadamente).

El SCL-90-R incluye 9 escalas, como por ejemplo las escalas de hostilidad, ansiedad o depresión, que no se corresponden directamente con los criterios diagnósticos sintomáticos del DSM-IV o la CIE-10. Además de las puntuaciones obtenidas en cada escala, el SCL-90-R calcula un Índice General de Gravedad a partir de la puntuación total de los distintos síntomas, así como una escala de Síntomas Positivos Totales (resultado de la suma total de síntomas puntuados como positivos) y un Índice de Malestar de Síntomas Positivos (cuya puntuación refleja la intensidad del malestar). La aplicación del cuestionario dura aproximadamente 12-15 minutos y su corrección está disponible en formato digital.



Se ha señalado con frecuencia que el SCL-90-R tiene una mayor fiabilidad en sus escalas generales que en las escalas dimensionales de psicopatología (Brophy, Norvell y Kiluk, 1988; Piersma, Boes, y Reume, 1994, en Bufka y Camp, 2011), y, por tanto, se ha concluido que tiene “poca utilidad como instrumento diagnóstico” (Bufka y Camp, 2011, p. 74). También se han encontrado estos mismos datos en la adaptación española (Martínez-Azumendi, Fernández-Gómez y Beitia-Fernández, 2001).

La correlación entre la escala de depresión del SCL-90 y la puntuación total del BDI es significativa (p. ej., Brophy, Norvell y Kiluk,). Además, especialmente en las escalas de depresión y ansiedad, los datos de un estudio en población alemana señalan que la eficacia diagnóstica del SCL-90 es aceptable (Schmitz, Kruse, Heckrath, Alberti y Tress, 1999).

Aunque el contenido de la escala de depresión no coincide completamente con los criterios diagnósticos del DSM-IV, una parte de los criterios están al menos parcialmente reflejados (p. ej., tristeza, ideación suicida, anhedonia, falta de energía, culpa y sentimientos de inutilidad). Por otro parte, aparecen algunos síntomas que no están recogidos por el DSM-IV (p. ej., pérdida del deseo sexual, preocupaciones o sentimientos de soledad).

#### **Adaptación española**

La adaptación española de este cuestionario (Derogatis, 2002, adaptación española J. L. González de Rivera, C. de las Cuevas, M. Rodríguez Abuin y F. Rodríguez Pulido) se realizó sobre una muestra de población general residente en Tenerife ( $N = 530$ ) y en una muestra psiquiátrica ambulatoria ( $N = 219$ ). Los baremos del cuestionario incluyen los valores normativos de la población clínica de ingreso, la población clínica ambulatoria, la población no clínica y la población

adolescente. Se encontraron buenos datos de fiabilidad de consistencia interna y de test retest para todas las escalas. En el caso de la escala de depresión se encontró una correlación entre las dos mitades de 0.87 y en el test-retest, con un intervalo de una semana, una correlación de 0.92. Se han encontrado mayores puntuaciones generales en aquellos pacientes con patología psiquiátrica que en personas de la población general, y puntuaciones aún superiores en una muestra de personas con trastorno depresivo mayor, lo que apoyaría en principio la validez de criterio de la adaptación española del SCL-90-R (de las Cuevas, 1991, en Derogatis, 2002). Sin embargo, aunque estas diferencias son especialmente significativas para la escala de depresión, parece que las diferencias entre estas 3 muestras son también significativas para todas las escalas, lo que cuestiona la validez discriminante de la escala de depresión del SCL-90-R. Por otro lado, no se ofrecen en el manual de la adaptación española del SCL-90-R índices de validez diagnóstica en relación con un criterio de diagnóstico externo para ninguna de sus escalas.

#### **3.3.3. Conclusiones acerca de los cuestionarios, escalas e inventarios para los trastornos depresivos**

De la literatura empírica se ha revisado en los dos anteriores epígrafes, se puede concluir que, en general, una buena parte de los estudios realizados para la valoración de la utilidad diagnóstica de los cuestionarios, escalas e inventarios más utilizados para evaluar los trastornos depresivos, sugieren que la mayoría de esos instrumentos, con sus limitaciones, son útiles para la detección de los trastornos depresivos, con aceptables o buenos datos de validez diagnóstica con respecto a las entrevistas estructuradas, aunque esos datos difieren mucho en función del tipo de muestra con la que se están utilizando. Obviamente, no es igual de fácil detectar un trastorno depresivo en una muestra de personas de la población general que en una muestra de pacientes médicos o en una muestras de pacientes psiquiátricos,

siendo, en este orden, mayor la dificultad y, por tanto, menores los índices de validez diagnóstica obtenidos. No obstante, la revisión realizada también indica que los datos publicados son en la mayoría de los casos insuficientes, tanto para las versiones originales de los instrumentos como, sobre todo, para sus adaptaciones en la población española.

Dependiendo del objetivo que se persiga a la hora de evaluar una persona, los instrumentos a utilizar serán diferentes. Si se tiene interés en que un instrumento sirva de cribado o *screening* o que sirva para valorar el cambio terapéutico o monitorizar la gravedad de los síntomas, las escalas de autoinforme como, por ejemplo, el BDI-II o el cuestionario PHQ-9 serán apropiadas por su fiabilidad, validez y sensibilidad. Sin embargo, incluso para esos objetivos, tales instrumentos tienen limitaciones, ya que, por ejemplo, para realizar una adecuada valoración de la sintomatología depresiva que permita profundizar en las particularidades de cada paciente, habría que considerar el uso de las entrevistas semiestructuradas, las cuales combinan las virtudes de las entrevistas clínicas en cuanto a la posibilidad de flexibilizar las preguntas, introducir preguntas nuevas y repreguntar para ampliar la información sobre la sintomatología del paciente y sus circunstancias, con las virtudes de las entrevistas estructuradas en cuanto ofrecen la posibilidad de realizar de forma más válida un diagnóstico diferencial, de valorar la comorbilidad de otros trastornos y, en principio, ofrecen una mayor fiabilidad en la evaluación que las entrevistas clínicas.

Si el objetivo de la evaluación es la valoración diagnóstica de los trastornos mentales, con frecuencia se debe recurrir a varias fuentes de información y a varios instrumentos para aumentar en lo posible su fiabilidad y validez, pero, en cualquier caso, parece que las entrevistas estructuradas y semiestructuradas son todavía, a falta de mejores opciones, el “patrón oro”, aunque los cuestionarios, escalas e

inventarios diseñados tanto para diagnosticar los trastornos depresivos como para cribar los trastornos depresivos pueden ser una alternativa válida en la medida en que tal validez se compruebe empíricamente en la población de interés, sin que existen suficientes estudios en la literatura científica que permitan dilucidar si a estos efectos los cuestionarios, escalas e inventarios de diagnóstico (p. ej., el DID o el PHQ-9) son mejores que los cuestionarios, escalas e inventarios de cribado (p. ej., el BDI-II, la HAMD, la SDS, etc.).

No obstante, los cuestionarios, escalas e inventarios de evaluación de los trastornos depresivos tienen un déficit en la investigación no sólo de sus propiedades de validez diagnóstica, sino en algunos casos, y sobre todo para las adaptaciones españolas, de las características psicométricas en general para la gran variedad de poblaciones en las que hoy se están utilizando. Una parte importante de estos problemas provienen de los siguientes hechos:

- Los estudios de fiabilidad y validez de algunos instrumentos se basan en versiones anteriores que no deben ser automáticamente asumidos cuando el instrumento es modificado.
- En ocasiones estos instrumentos tienen como única validación los estudios originales de los propios autores que lo han desarrollado, siendo estos estudios habitualmente más favorables que los independientes (Craig, 2005a).
- Los estudios de validez publicados ofrecen con frecuencia datos de eficacia diagnóstica muy limitados, que no proporcionan información suficiente sobre las características del instrumento.
- Los estudios de adaptación a otros idiomas o poblaciones no se realizan en profundidad por parte de los adaptadores. Además, no son frecuentes los estudios de replicación por parte de grupos de investigación independientes.

**Capítulo 4. EL INVENTARIO  
CLÍNICO MULTIAXIAL DE  
MILLON (MCMI) COMO  
INSTRUMENTO PARA EL  
DIAGNÓSTICO DE LOS  
TRASTORNOS DEPRESIVOS**



#### **4.1. Características generales del Inventario Clínico Multiaxial de Millon (MCMI)**

El MCMI es un instrumento con un gran reconocimiento en el ámbito de la psicopatología desde hace casi 30 años. Las tres versiones del inventario (MCMI-I, MCMI-II y MCMI-III; Millon, 1983, 1987, 1994; Millon, Davis y Millon, 1997; Millon, Millon, Davis y Grossman, 2009) se caracterizan por recoger información acerca de una gran variedad de trastornos de personalidad y trastornos clínicos. Este inventario está diseñado para ser usado en la planificación del tratamiento y la psicoterapia, en la atención a pacientes con abuso de sustancias, en el ámbito forense, en neuropsicología, en ámbitos penitenciarios y en consejo de pareja (Millon et al., 2007). Según el manual del MCMI-II (Millon, 1998, p. 16-17):

“El principal objetivo del MCMI-II es proporcionar información a los clínicos (psicólogos, psiquiatras, consejeros clínicos, trabajadores sociales, enfermeras) en tareas de evaluación y tratamiento de personas con dificultades emocionales e interpersonales. El perfil de las puntuaciones de las escalas del MCMI-II puede servir para identificar qué grupo de personas con dificultades emocionales y personales puede requerir una evaluación más intensa o atención profesional”.

Las características básicas del instrumento se han conservado desde su creación (véase la Tabla 4.1). El MCMI se compone de 175 ítems con respuestas de verdadero o falso. Está formado por escalas clínicas de personalidad, escalas de personalidad de patología grave, escalas de síndromes clínicos y escalas de síndromes graves (véase el Anexo 5). A pesar de que recoge un número elevado de escalas, su tiempo de aplicación es relativamente breve, aproximadamente 20-30 minutos. El MCMI ha sido desarrollado para su aplicación en personas mayores de 18 años pertenecientes a la población clínica. Millon (1998, p. 17) señala que “el

MCMI-II no es un instrumento de evaluación de la personalidad de poblaciones consideradas *normales* ni tiene otros propósitos que no sean la exploración diagnóstica o la evaluación clínica”.

**Tabla 4.1.** Características básicas del MCMI-I, MCMI-II y MCMI-III

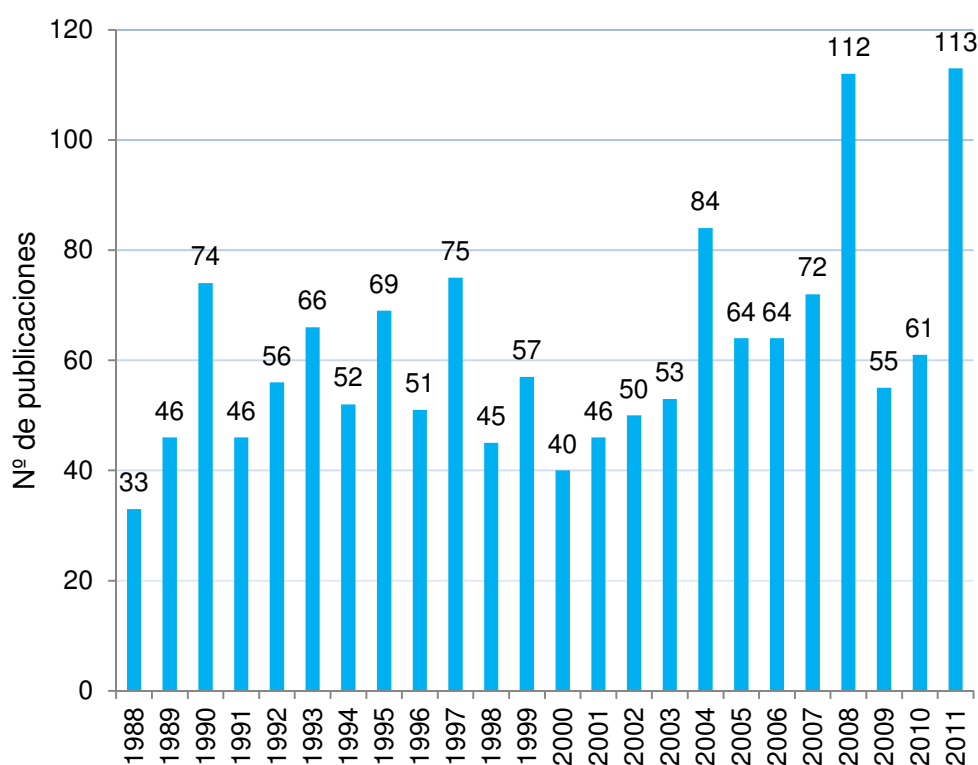
Características	MCMI-I	MCMI-II	MCMI-III
Número de ítems	175	175	175
Escalas de personalidad (incluidas las de patología grave)	11	13	14
Escalas de síndromes clínicos (incluidas las de síndromes graves)	9	9	10
Ámbito de aplicación	Adultos	Adultos	Adultos

El MCMI es uno de los instrumentos psicopatológicos más utilizados tanto en investigación como en la práctica clínica diaria. Una búsqueda en la base de datos bibliográficos PsycINFO mediante las palabras “MCMI” o “Millon Clinical Multiaxial Inventory”, en los campos “resumen”, “pruebas y medidas” o “descriptores”, encontró 1484 referencias (de las cuales 1005 se referían a artículos de revistas) en las cuales se había utilizado o analizado el MCMI. El análisis se realizó para el período comprendido entre el 1 de enero de 1988, año de aparición de la primera versión del MCMI, y el 31 de diciembre de 2011. En la Figura 4.1 se recogen los trabajos encontrados en función del año de su publicación. Como puede observarse en dicha figura, en los últimos 24 años, el uso del MCMI ha crecido de manera importante y estable. De hecho, en términos del número de publicaciones recogidas en la base de datos PsycINFO, su utilización en los últimos años es comparable a la de instrumentos psicopatológicos tan frecuentes y asentados como el test de Rorschach, aunque todavía se encuentra lejos de los

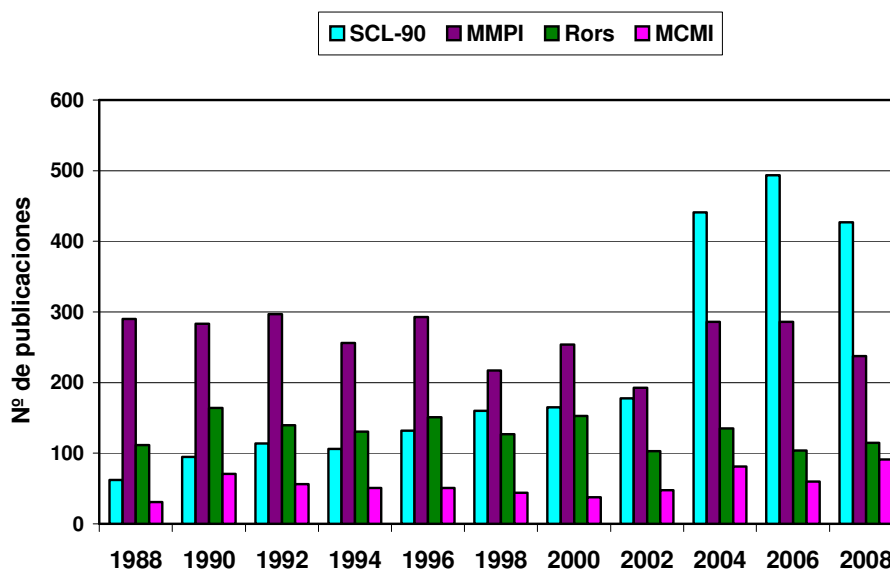


instrumentos psicopatológicos actualmente más populares: el Inventario Multifásico de Personalidad de Minnesota (MMPI) y el Cuestionario de 90 Síntomas (SCL-90) (véase la Figura 4.2). En España, el MCMI también ha ido ganando popularidad entre los investigadores y profesionales, de forma que en los últimos años la adaptación española del MCMI-II ha sido utilizada en más de una treintena de investigaciones, y, en casi la mitad de ellas, el MCMI-II fue utilizado para realizar diagnósticos (Sanz, 2007).

**Figura 4.1.** Búsqueda en los campos “resumen”, “descriptores” y “pruebas y medidas” de PsycINFO de “MCMI” o “Millon Clinical Multiaxial Inventory”.



**Figura 4.2.** Búsqueda en los campos “resumen”, “descriptores” y “pruebas y medidas” de PsycINFO de “MMPI” o “Minnesota Multiphasic Personality Inventory”, “MCMI” o “Millon Clinical Multiaxial Inventory”, “Rorschach” y “SCL-90” o “Symptom Checklist-90”.



*Nota.* SCL-90: Cuestionario de los 90 Síntomas; MMPI: Inventario Multifásico de Personalidad de Minnesota; Rors: Rorschach; MCMI: Inventario Clínico Multiaxial de Millon.

La construcción de este instrumento está muy influida por el modelo dimensional de psicopatología defendido por su principal autor, Theodore Millon. Desde el nacimiento de la primera versión del MCMI, Millon y sus colaboradores han tratado de aunar en este instrumento varios aspectos: el propio modelo teórico de la personalidad y la psicopatología de Millon, una metodología científico/práctica y los criterios diagnósticos de los trastornos psicológicos recogidos en las distintas versiones del DSM. Para ello, el modelo dimensional de Millon ha debido ser encajado con los modelos categoriales. Sin embargo, la relación entre la teoría de Millon y el contenido del MCMI es mucho más fuerte en el caso de las escalas de trastornos de personalidad, las relacionadas con el Eje II del DSM, que en el caso de las escalas de síndromes clínicos, las relacionadas con el Eje I del DSM, las cuales se derivan más claramente del propio DSM. En este sentido, la versión española del manual del MCMI-III señala que: “Aunque las escalas del Eje I no

están derivadas explícitamente de la teoría, no obstante se han elaborado en relación con su marco teórico” (Millon et al., 2007, p. 18). Puesto que la presente tesis doctoral está centrada en el análisis de dos escalas de síndromes clínicos del MCMI-II y del MCMI-III, las escalas de Depresión Mayor y de Distimia, en aras de la brevedad se remite al lector a las numerosas publicaciones en español destinadas al análisis en profundidad de la teoría de Millon para una descripción de ésta (p. ej., Cardenal, Sánchez y Ortiz-Tallo, 2007a; Millon, 1974, 1976; Millon y Davis, 1998), mientras que más adelante se abordará con detalle la relación de las escalas de Depresión Mayor y Distimia del MCMI-II y del MCMI-III con la última versión publicada del DSM, el DSM-IV.

El MCMI es un proyecto desarrollado durante las últimas décadas que aún está en continuo proceso de crítica y cambio. Es un inventario complejo en su estructura, complicado de corregir (es imprescindible una corrección automatizada) y difícil de interpretar en la práctica clínica y en la investigación, por lo que, además de sus propios manuales, se dispone de un buen número de publicaciones para su correcta utilización (Cardenal, Sánchez y Ortiz-Tallo, 2007b; Craig, 2005b; Millon y Bloom, 2008). Sin embargo, algunos aspectos de este instrumento son controvertidos como, por ejemplo, su validez de contenido, sus baremos de corrección y su validez diagnóstica (Hsu, 2005), cuestiones todas ellas que se desarrollarán más adelante.

Uno de los objetivos del instrumento es el de proporcionar una herramienta de evaluación para un amplio espectro de trastornos psicológicos. En este sentido, Millon (1998, p. 160) afirma:

“Los resultados del MCMI-II pueden emplearse de distintas maneras con propósitos clínicos. El uso principal es la interpretación de un perfil mediante un proceso de inferencia de los aspectos más importantes de la

configuración de las 22 escalas clínicas. Otro uso importante es la toma de decisiones para identificar el diagnóstico específico que mejor se aplica al paciente”.

Según el manual del MCMI-III (Millon et al., 2007, pp. 20 y 21):

“...el MCMI intenta diagnosticar los porcentajes de los pacientes que realmente presentan trastornos en los ámbitos diagnósticos. Estos datos no sólo facilitan una base para seleccionar los puntos de corte óptimos de diagnóstico diferencial, sino que también garantizan que la frecuencia de los diagnósticos generados por el MCMI-III y de los patrones de los perfiles serán comparables con las tasas representativas de prevalencia clínica”.

Como se ha comentado antes, el MCMI se compone de una gran variedad de escalas, con distribuciones muy desiguales y que evalúan trastornos con distintas prevalencias en la población. Este hecho genera dificultades para cumplir con el objetivo diagnóstico que persigue el inventario, por lo que Millon y su equipo diseñaron un sistema para convertir las puntuaciones directas de las escalas en puntuaciones tasa base (TB) que tratan de resolver tales dificultades.

#### **4.2. Las puntuaciones tasa base (TB) o de prevalencia como herramienta diagnóstica**

Para la interpretación de las puntuaciones de los sujetos en los cuestionarios, escalas e inventarios psicopatológicos existen varios procedimientos. Con frecuencia, los evaluadores examinan las puntuaciones directas y fijan un punto de corte a partir del cual se asume la presencia de un trastorno en el sujeto. En el caso de los cuestionarios que miden la gravedad de la sintomatología depresiva, muchos de ellos suelen proporcionar distintos puntos de corte para

asignar a los pacientes a una categoría de gravedad de la depresión, puntos de corte que se pueden calcular a su vez mediante diversos procedimientos como, por ejemplo, analizando las puntuaciones que mejor discriminan entre distintas categorías de gravedad definidas por un criterio externo, en particular los juicios de gravedad sindrómica proporcionados por uno o varios clínicos (este es el caso, por ejemplo, del BDI-II; Beck et al., 1996). En otras ocasiones, el método elegido para interpretar la puntuación de un sujeto determinado es transformar su puntuación en un valor que le sitúa en una determinada posición con respecto a un grupo de referencia, generalmente mediante el cálculo de las puntuaciones típicas. Este método permite comparar al sujeto con respecto a la media y la desviación típica de una muestra, pero sin tener en cuenta la prevalencia de los trastornos en la población, lo que provoca que existan errores sistemáticos debido a que los puntos de corte no suelen permitir la detección del mismo porcentaje de sujetos que tienen una patología en una población (Millon, 1997).

Con el objetivo de subsanar estas dificultades, el MCMI utiliza un método alternativo que tiene en cuenta que el sujeto pertenece a una población en la que existe una prevalencia determinada de los distintos trastornos psicopatológicos. En los estudios para la construcción de los baremos, los pacientes son clasificados en función de los diagnósticos que realizan los clínicos. Además, si se detecta más de un diagnóstico se debe señalar cuál es considerado como más prominente en cada paciente. Por ejemplo, en el estudio realizado para la adaptación española del MCMI-III (Millon et al., 2007, adaptación española de V. Cardenal y M. P. Sánchez), se estimó que existía en la población clínica española un 8% de pacientes con distimia como trastorno más prominente y un 21% de pacientes en los que se detectó dicho trastorno aunque no fuera el más prominente. Por lo tanto, si se aplica el MCMI-III en otras muestras de la población clínica española, se deberían detectar esos mismos porcentajes de prevalencia del trastorno distímico. Para

lograr este objetivo, el procedimiento de conversión de puntuaciones directas a puntuaciones de tasa base (TB) del MCMI implica que se asigne de forma arbitraria una puntuación TB 85 a la puntuación directa que deja por encima el mismo porcentaje de pacientes que tienen distimia como trastorno más prominente, en el ejemplo anterior un 8%. La puntuación TB 75 indicaría el punto de corte que deja por encima el mismo porcentaje que la prevalencia del trastorno, independientemente de que sea el más prominente o no, en el ejemplo anterior un 21%. La puntuación TB 60 equivaldría a la media de las puntuaciones directas de la escala en la población clínica, mientras que la TB 35 correspondería a la media de las puntuaciones directas de la escala en la población no clínica. A partir de estas asignaciones estándar y de la asignación de la puntuación directa de 0 a la puntuación TB de 0 y de la puntuación directa máxima de la muestra a la puntuación TB de 115, el resto de asignaciones entre las restantes puntuaciones directas y las correspondientes puntuaciones TB se realizan por interpolación lineal, aunque tanto los autores de la versión original como los autores de las adaptaciones españolas señalan en sus correspondientes manuales que para poder completar la asignación de puntuaciones TB se realizaron otros ajustes y modificaciones cuyos procedimientos y justificaciones no quedan descritos exactamente en dichos manuales.

El sistema de puntuaciones TB surge como respuesta a un problema específico del MCMI y, probablemente, de algunos otros cuestionarios, escalas e inventarios psicopatológicos. Se ha comprobado que en algunas escalas del MCMI las puntuaciones se distribuyen de forma asimétrica, de manera que muchos sujetos tienen puntuaciones bajas y unos pocos tienen puntuaciones altas. Por tanto, no se trata de unas puntuaciones con una distribución normal y las frecuencias son muy diferentes entre los trastornos. Las puntuaciones altas en algunas escalas son mucho más frecuentes que en otras, lo que complica la

interpretación del inventario. No es deseable que se detecten, por ejemplo, el mismo número de pacientes con sintomatología psicótica que con trastornos depresivos. Además, Millon (1998, p. 73) comenta que:

“Otra razón para emplear los datos reales de prevalencia es que el propósito principal de un instrumento clínico no es ubicar la posición relativa de un paciente en una distribución de frecuencias, sino calcular la probabilidad que el paciente tiene de formar parte de una entidad diagnóstica particular”.

Es importante destacar que los baremos de corrección que utilizan este sistema deberían ser diferentes en cada población, porque existen diferencias en la prevalencia y en los perfiles de respuesta de los sujetos en función de su población de origen (Ayuso-Mateos et al., 2001; Demyttenaere et al., 2004; Kessler et al., 2007a, 2010).

Sin embargo, ni el hecho de que el propósito de las puntuaciones TB sea maximizar la probabilidad de alcanzar un diagnóstico válido ni el propio mecanismo de obtención de las puntuaciones TB garantizan automáticamente que los puntos de corte asignados a las TB 75 y 85 tengan buenas propiedades psicométricas en relación con su validez diagnóstica. Por ejemplo, el hecho de que una persona tenga una puntuación menor al punto de corte TB 75 no significa automáticamente que dicha persona no cumpla los criterios para el diagnóstico correspondiente. Además, una persona puede puntuar más alto que el punto de corte correspondiente a la prevalencia del trastorno, sin cumplir los criterios diagnósticos del mismo. De hecho, en el pasado ya se han encontrado ejemplos en los que los puntos de cortes de las escalas del MCMI-I desarrollados bajo este sistema de TB han sobrediagnosticado la presencia de sujetos con depresión mayor con respecto al diagnóstico realizado por clínicos (p. ej., Wetzler, 1990).

En resumen, el sistema de conversión de puntuaciones directas en puntuaciones TB está diseñado con el objetivo de realizar a través del MCMI un diagnóstico clínico con buenas cualidades psicométricas. Este sistema ha sido puesto en duda debido a que es posible que existan otros métodos específicamente diseñados para optimizar la validez diagnóstica de los puntos de corte (Hsu, 2002). Por tanto, sería un error dar por hecho que las puntuaciones de TB 75 y TB 85 se corresponden automáticamente con las puntuaciones que logran la mejor validez diagnóstica, sino que esta cuestión debería ser objeto de investigación empírica.

#### **4.3. Características y versiones del Inventario Clínico Multiaxial de Millon: MCMI-I, MCMI-II y MCMI-III**

##### **4.3.1. Primera versión del Inventario Clínico Multiaxial de Millon (MCMI-I)**

Para la creación del MCMI-I se partió de un banco de ítems (Millon, 1983). Mediante un método racional se seleccionaron aquellos ítems que mejor ejemplificaban las distintas características de cada una de las escalas que componen el cuestionario en base a la teoría de Millon. En las sucesivas versiones (MCMI-II y MCMI-III; Millon, 1987, 1994; Millon et al., 1997) se han añadido elementos que se corresponden con los criterios diagnósticos de los trastornos de las distintas ediciones del manual DSM.

El MCMI-I (Millon, 1983) se compone de 11 escalas de trastornos de personalidad (Esquizoide, Evitativo, Dependiente, Histriónico, Narcisista, Antisocial, Compulsivo, Pasivo-Agresivo, Esquizotípico, Límite y Paranoide) y 9 escalas sobre síndromes clínicos (Ansiedad, Somatomorfo, Hipomaniaco, Distímico, Abuso de Alcohol, Abuso de Drogas, Pensamiento Psicótico, Depresión Psicótica y Trastorno Delirante).



En las diversas revisiones realizadas sobre los estudios del MCMI-I se ha señalado la falta de concordancia entre los diagnósticos realizados por los clínicos (según el manual DSM-III) y los perfiles obtenidos mediante el inventario (Inch y Crossley, 1993). La revisión más completa sobre esta versión del instrumento analizaba sus propiedades principales y llegaba a las siguientes conclusiones sobre el mismo (Wetzler, 1990):

- En el inventario, siguiendo la teoría de Millon, un alto porcentaje de ítems forman parte de varias escalas al mismo tiempo, de manera que las correlaciones entre escalas llegan a ser excesivamente altas, siendo en algunos casos cercanas a 1. De hecho “las relaciones entre las escalas parecen ser un artefacto más que el resultado intrínseco de las relaciones conceptuales entre las dimensiones” (Wetzler, 1990, p. 447).
- La fiabilidad del test proporcionada en el manual es alta (coeficiente de Kuder-Richarson medio = 0.88; Millon, 1983), indicando que las escalas del MCMI tienen buena consistencia interna.
- Los datos parecen apoyar una mayor estabilidad en las puntuaciones de las escalas de trastornos de personalidad, mientras que las puntuaciones en las escalas de síndromes clínicos son más dependientes del estado momentáneo del sujeto. En este sentido, las escalas del MCMI-I que miden trastornos del estado de ánimo demostraron buena sensibilidad a la situación temporal de los pacientes (p. ej., Choca, Bresolin, Okonek y Ostrow, 1988).

Según Wetzler (1990, p. 453), para el MCMI-I Millon escogió:

“un enfoque conservador, tratando de maximizar la especificidad (el porcentaje de pacientes sin el trastorno que eran correctamente no diagnosticados por el test) y el poder predictivo negativo (el porcentaje de

pacientes diagnosticados por el test como ausencia de trastorno que efectivamente no cumplen los criterios del trastorno)”.

Se publicaron en el manual los datos de eficacia diagnóstica del MCMI-I con respecto a los diagnósticos realizados a partir de los criterios narrativos derivados de la teoría de Millon (Millon, 1983). Salvo excepciones, los datos fueron positivos. Por ejemplo, el poder predictivo positivo fue mayor de 0.70 para la escala de Distimia. Aunque en particular algunas escalas mostraban datos inferiores a 0.50, como era el caso de la escala de Depresión Psicótica.

En la corrección del MCMI-I se sugerían diagnósticos basado en el DSM-III, pero su eficacia diagnóstica fue controvertida (Wetzler, 1990). Existen pocos estudios realizados con posterioridad a la publicación del manual y sus resultados son muy dispares para cada escala. En el caso de la escala de Depresión Psicótica, se encontraron datos muy malos con respecto al diagnóstico de trastorno depresivo mayor, identificándose tan solo un 4% de pacientes con dicho diagnóstico con  $TB \geq 75$  (Wetzler, Kahn, Strauman y Dubro, 1989, en Wetzler, 1990) y un 10-15% con  $TB \geq 85$  (Choca et al., 1988; Goldberg, Shaw y Segal, 1987, en Wetzler, 1990). Respecto a la escala de Distimia, se encontraron datos significativamente mejores (Wetzler, 1990). Los resultados de este estudio empírico señalaban, según Wetzler, (1990, p. 454), que “la escala de Distimia parece detectar aspectos de la depresión que no tienen relación con el estado actual del paciente, porque identifica un 50% de los pacientes con estado eutímico, que han tenido en el pasado una historia de depresión”.

Otro estudio en el que se estudiaron las propiedades diagnósticas de las escalas de Depresión Psicótica y Distimia del MCMI-I fue realizado en una muestra de 116 sujetos psiquiátricos (Wetzler, Hahn, Strauman y Dubro, 1989). Sus resultados mostraban que la sensibilidad era alarmantemente baja en el caso de la

escala de Depresión Psicótica (véase la Tabla 4.2), por lo que los autores recomendaban que: “En ninguna circunstancia se debería indicar la escala de depresión psicótica como instrumento de evaluación según los datos aportados” (Wetzler et al., 1989, p. 29).

**Tabla 4.2.** Estadísticos de validez diagnóstica de las escalas de Distimia y Depresión Psicótica del MCMI-I del estudio de Wetzler et al. (1989)

	<b>Sensibilidad</b>	<b>Especificidad</b>	<b>PPP</b>	<b>PPN</b>	<b>Prevalencia</b>
Distimia	71	70	77	63	41
Depresión Psicótica	04	97	59	50	41

*Nota.* PPP = poder predictivo positivo; PPN = poder predictivo negativo.

Wetzler (1990, p. 462) señalaba que la principal limitación del MCMI-I estriba en su falta de capacidad para realizar diagnósticos válidos, sobre todo para algunas de sus escalas:

“En cualquier caso, mi mayor preocupación sobre el MCMI es su aparente compatibilidad con el DSM-III. Los usuarios ingenuos del test pueden creer que las escalas con etiquetas del DSM-III y los informes computerizados con recomendaciones sobre diagnósticos DSM-III, son realmente indicativos de síndromes del DSM-III... aunque algunas escalas son compatibles con el DSM, otras no lo son”.

#### 4.3.2. Inventario Clínico Multiaxial de Millon-II (MCMI-II)

Con el objetivo de resolver algunos de los problemas señalados en el epígrafe anterior, se publicó el MCMI-II (Millon, 1987, 1994; Millon, 1997) con cambios realizados para acercar el contenido de las escalas al DSM-III-R. Millon (1998, p. 13) afirmaba en el manual del MCMI-II que:

“Como su predecesor, sin embargo, el MCMI-II sigue sin ser una herramienta diagnóstica perfecta. Ni refleja exactamente el fundamento

teórico de los constructos clínicos ni refleja todas las facetas de los trastornos sindrómicos adoptados por el DSM-III-R, al que pretende estar coordinado. A pesar de las restricciones de validez impuestas por las limitaciones de los autoinformes, las estrechas fronteras de la tecnología psicométrica y el escaso grado de consenso sobre el conocimiento diagnóstico, se dieron todos los pasos para aumentar al máximo la concordancia del MCMI-II con la teoría en la que se basa y con el sistema oficial de clasificación”.

En esta versión del cuestionario se incorporaron cambios con respecto a su antecesor (Millon, 1987; Inch y Crossley, 1993):

- Se realizaron cambios en 45 ítems del cuestionario. En el caso de los síndromes clínicos, las modificaciones se realizaron con el DSM-III-R (APA, 1987) de referencia. Los resultados de las investigaciones sobre la validez del MCMI-I llevaron a introducir cambios con el objetivo de mejorar su “exactitud diagnóstica” (Millon, 1997). Una de las escalas que más modificaciones sufrió fue precisamente la escala de Depresión Psicótica, que pasó a llamarse de Depresión Mayor.
- Se incluyó un nuevo sistema de corrección que asignaba diferentes pesos a los diferentes ítems. Este sistema ha sido defendido por sus autores para dar más relevancia a los ítems que son “más prototípicos de un diagnóstico” (en cuyo caso se valora con 3 puntos) o que forman parte de más de una escala (valorados con 2 puntos). El resto de ítems se valora con 1 punto. “El propósito era reducir las correlaciones extremadamente altas entre las escalas, causadas por el hecho de que existe un considerable grado de solapamiento de ítems” (Streiner, Golberg y Miller, 1993, p. 472). Estos autores encontraron que la correlación entre la puntuación directa de las escalas utilizando el peso de los

ítems y sin tenerlo en cuenta era superior a 0.90 para todas las escalas, por lo que concluían que la corrección con diferente ponderación de los ítems provocaba que el instrumento se volviera menos accesible para los clínicos, sin que se aumentasen sus cualidades psicométricas. En la línea de estos datos, una investigación realizada en una muestra de 356 sujetos mostró una correlación cercana a 1 entre las escalas modificadas por el peso de los ítems y las que no fueron modificadas, por lo que pone también en duda la utilidad de los pesos de los ítems (Retzlaff, Sheehan y Lorr, 1990).

- Se modificaron las escalas de validez, siendo accesibles para los evaluadores (Índice de Validez, Deseabilidad Social y Devaluación). Estas escalas tienen como objetivo mejorar la sensibilidad y la especificidad del cuestionario mediante la identificación de patrones de respuesta marcados por la aleatoriedad, la deseabilidad social o la voluntad de los sujetos para aparecer como más o menos afectados emocionalmente (Millon, 1987). En su construcción se siguió un método racional validado empíricamente en pequeñas muestras de estudiantes y pacientes (Millon, 1987).

En la publicación original del MCMI-II, para desarrollar los baremos de transformación de puntuaciones directas en puntuaciones TB se analizaron los datos de 1292 pacientes, evaluados por más de 600 clínicos. Se asignaron diagnósticos a los pacientes y, en el caso de que hubiera más de un diagnóstico, se señaló cuál era más prominente en cada caso. Tras calcular la prevalencia de los distintos trastornos se pudieron determinar los puntos de corte TB 75 (presencia de trastorno) y TB 85 (trastorno más prominente). En el manual original del MCMI-II no se explicita si estos diagnósticos se realizaron con la ayuda de entrevistas diagnósticas, la puntuación del propio instrumento o cualquier otro elemento de ayuda durante la evaluación. Tampoco es posible conocer el número de sesiones

realizadas hasta el momento del diagnóstico. Por otro lado, en el manual original del MCMI-II se señala que los datos psicométricos de eficacia diagnóstica de sus escalas son mejores que los de las escalas de su predecesor (Millon, 1987).

Los coeficientes de fiabilidad aportados por el manual del MCMI-II muestran niveles aceptables de fiabilidad test-retest (entre 3 y 5 semanas de tiempo entre la primera y la segunda aplicación), teniendo en cuenta que la muestra se encontraba con frecuencia en un proceso terapéutico, siendo en la mayor parte de los casos superiores a 0.5 (Inch y Crossley, 1993).

### **Datos sobre la validez diagnóstica del MCMI-II**

Una vez calculadas la correspondencia entre las puntuaciones directas y las puntuaciones TB, se utilizó una muestra independiente de 703 pacientes para comprobar la validez externa de los datos encontrados y para comprobar sus propiedades psicométricas para el diagnóstico de las puntuaciones TB  $\geq 75$  y TB  $\geq 85$  (Millon, 2002). Las evaluaciones diagnósticas incluyeron 6 o más entrevistas con los pacientes y se realizaron sin que los evaluadores conocieran los resultados del MCMI-II. Como ayuda para la asignación del diagnóstico los terapeutas tenían un listado con las descripciones de los trastornos más frecuentes con los criterios DSM-III-R. Los datos sobre validez diagnóstica encontrados fueron por lo general positivos (véanse la Tabla 4.3. y la Tabla 4.4), con una especificidad por encima del 0.83 y una sensibilidad que oscilaba entre el 0.34 y el 0.78, con una media superior al 0.60 (Millon, 2002). Se puede observar, por tanto, como efectivamente el MCMI-II es un instrumento que prioriza la especificidad por encima de la sensibilidad.

**Tabla 4.3.** Índices de eficacia diagnóstica de las escalas del MCMI-II (punto de corte de TB > 75) para el diagnóstico de los correspondientes síndromes clínicos del Eje I del DSM-III-R (Millon, 2002)

Escala	Prevalencia	Sens	Espe	PPP	PPN	PDT
Somatoforme	13	57	92	53	92	88
Hipomanía	4	42	99	55	98	97
Distimia	46	81	83	80	84	82
Abuso de alcohol	15	87	99	92	98	97
Abuso de drogas	13	72	98	82	96	94
Pensamiento psicótico	4	37	99	850	98	96
Depresión mayor	14	68	96	73	95	92
Trastorno delirante	5	34	99	61	97	96

*Nota.* Sens = Sensibilidad; Espe = Especificidad; PPP = poder predictivo positivo; PPN = poder predictivo negativo; PDT = poder diagnóstico total.

**Tabla 4.4.** Índices de eficacia diagnóstica de las escalas del MCMI-II (punto de corte de TB ≥ 85) para el diagnóstico de los correspondientes síndromes clínicos del Eje I del DSM-III-R (Millon, 2002)

Escala	Prevalencia	Sensibilidad	Especificidad	PPP	PPN	PDT
Ansiedad	23	69	93	74	91	88
Somatoforme	5	43	95	33	97	92
Hipomanía	3	40	99	53	98	97
Distimia	29	76	88	72	90	84
Abuso de alcohol	8	79	99	88	98	97
Abuso de drogas	9	62	98	78	97	95
Pensamiento psicótico	2	20	99	30	98	97
Depresión mayor	6	47	97	50	97	94
Trastorno delirante	1	25	99	40	98	99

*Nota.* PPP = poder predictivo positivo; PPN = poder predictivo negativo; PDT = poder diagnóstico total.

#### 4.3.3. Inventario Clínico Multiaxial de Millon-III

Tras la publicación del DSM-IV, se desarrolló la última versión del MCMI publicada hasta la fecha, el MCMI-III (Millon, 1994; Millon et al., 1997). En esta ocasión se volvieron a realizar cambios muy significativos con respecto a su antecesor (Rogers, Salekin y Kenneth, 1999; Millon y Davis, 1997):

- Se añadieron dos nuevas escalas, trastorno de la personalidad depresivo y trastorno de estrés postraumático.
- Se modificaron más de la mitad de los ítems del cuestionario para acercarlos a los nuevos criterios diagnósticos del DSM-IV y mejorar su validez.
- Se redujo el solapamiento de ítems entre las escalas y se disminuyó el número de ítems que componía cada escala.
- Se realizaron cambios en el peso otorgado a los ítems de cada escala (se asignaron 2 puntos a los ítems considerados como prototípicos de cada escala, mientras que al resto de elementos se les asignó 1 punto) así como en las escalas de modificación.

En resumen, en la Tabla 4.5 se recogen los principales cambios que se han producido a lo largo de las tres versiones publicadas hasta la fecha del MCMI.

**Tabla 4.5.** Cambios que se han producido en las diferentes versiones del MCMI: MCMI-I, MCMI-II y MCMI-III (Rogers, Salekin y Kenneth, 1999)

Versión	Ítems nuevos (%)	Nuevas escalas	Nuevos pesos	Solapamiento de ítems	Longitud de las escalas (media)
MCMI	NA	NA	NA	4.21 escalas <sup>a</sup>	36.80 ítems
MCMI-II	45 (27.5%)	5	Si	5.44 escalas <sup>a</sup>	38.08 ítems
MCMI-III	95 (54.3%)	2	Si	2.50 escalas <sup>a</sup>	16.80 ítems

Nota. NA = no aplicable; <sup>a</sup> = Número de escalas a las que pertenece cada ítem de media.



### **Primera edición del manual del MCMI-III (Millon, 1994)**

Un procedimiento similar al utilizado con el MCMI-II se llevó a cabo para la elaboración de las puntuaciones TB del MCMI-III en la primera versión del manual (Millon, 1994).

Al analizar la información psicométrica proporcionada en dicho manual, lo primero que llama la atención es que los datos sobre la sensibilidad y especificidad de las escalas del MCMI-III fueron excepcionalmente bajos, inferiores a los encontrados en el MCMI-II. Las diferencias entre la segunda versión del test y la tercera son considerables, aunque muchos de estos cambios conservaban en esencia el contenido del test y los datos disponibles sobre la validez convergente de las escalas eran positivos cuando se utilizaba para ello cuestionarios como el STAI, el SCL-90-R o el MMPI. Por tanto, estos cambios no parecían justificar la reducción en la validez diagnóstica del MCMI-III en comparación a la versión anterior (Retzlaff, 1996).

Estudiando el procedimiento seguido por los investigadores surgen muchas dudas sobre la metodología utilizada en el estudio de validez diagnóstica de las escalas del MCMI-III. Según se describe en el manual los clínicos debían emitir el diagnóstico de sus pacientes tras haber aplicado la versión experimental del MCMI-III, es decir, mientras se estaba decidiendo los ítems definitivos del cuestionario, sin que ellos estuvieran al tanto de que los datos serían utilizados para los estudios de validez diagnóstica. Además, la mayor parte de los psicólogos implicados en la investigación conocían a sus pacientes de una sola sesión antes de emitir el juicio diagnóstico y no contaban con la ayuda de entrevistas estructuradas o la posibilidad de leer historiales previos (Millon, 1994). Estas dificultades podrían explicar la existencia de una subestimación de los datos de validez diagnóstica del MCMI-III en la primera versión de su manual.

En este primer manual del MCMI-III (Millon, 1994) se aportaban los datos referidos a la prevalencia, la sensibilidad y la especificidad. En una publicación posterior, Rezlaff (1996) calculó en base a esos datos el poder predictivo positivo y el poder predictivo negativo. Como cabría esperar, estos índices fueron también inferiores a lo esperado, sobretodo en cuanto al poder predictivo positivo (Rezlaff, 1996).

### **Segunda edición del manual del MCMI-III (Millon et al., 1997)**

Ante las dudas planteadas sobre los procedimientos seguidos en los estudios de validez recogidos en el primer manual del MCMI-III y ante la escasez de datos psicométricos publicados para la baremación de las escalas, se realizaron importantes esfuerzos por aportar un análisis más completo de las escalas del MCMI-III en una nueva edición de su manual.

En esta segunda edición del manual del MCMI-III (Millon et al., 1997), los datos de validez diagnóstica fueron significativamente más positivos. Para la realización del estudio se solicitó a los clínicos que emitieran su juicio diagnóstico cuando conocieran lo suficiente a sus pacientes (al menos tras realizar tres sesiones de evaluación de una hora), utilizando para ello los instrumentos que fueran precisos, excluyendo el MCMI. Además en este estudio, con el objetivo de facilitar que los psicólogos tuvieran la información sobre los pacientes disponibles con facilidad, se les presentó el listado de síndromes que debían tener en cuenta para realizar el diagnóstico. En este listado, los trastornos coincidían con las escalas del MCMI-III en el nombre y en el orden en el que se presentaban. Así mismo, se incluía una ficha en la que los terapeutas debían registrar tanto las puntuaciones TB del MCMI-III como los diagnósticos clínicos, pero se recomendaba a los evaluadores que no rellenaran las puntuaciones del test hasta que no se realizaran los diagnósticos.

Con el objetivo de recoger la máxima información posible, se dieron instrucciones a los psicólogos para que diagnosticaran también a los pacientes de los que ya supieran sus puntuaciones del MCMI, siempre que no las recordaran nítidamente. Existen argumentos para suponer que no se controló lo suficiente esta importante variable contaminadora. Se puede considerar que los psicólogos clínicos que participaban en la investigación, interesados por lo general en las puntuaciones del MCMI para comprender las características de sus pacientes, con facilidad se vieron influidos por estas puntuaciones en la conceptualización que tenían de ellos. Más aún cuando, como se ha comentado, los evaluadores debían utilizar la misma hoja para indicar el diagnóstico de su paciente y escribir sus puntuaciones en las distintas escalas del MCMI-III (Grove y Vrieze, 2009).

Estos aspectos pudieron incidir en que en este estudio, al contrario de lo que ocurrió en el estudio de la primera edición del manual original, se pudiera sobreestimar la validez diagnóstica del inventario. En la Tabla 4.6 se presentan los índices de sensibilidad y poder predictivo positivo recogidos en la segunda edición del manual original del MCMI-III (tal y como aparecen publicados en el manual de la versión española del MCMI-III; Millon et al., 2007, adaptación española V. Cardenal y M. P. Sánchez).

**Tabla 4.6.** Índices de eficacia diagnóstica de las escalas del MCMI-III (puntuación de corte de TB  $\geq$  85) para los correspondientes síndromes clínicos del Eje I del DSM-IV (Millon et al., 2007)

Escala	Prevalencia		Sensibilidad		PPP	
	MCMI-III <sup>a</sup>	MCMI-III <sup>b</sup>	MCMI-III <sup>a</sup>	MCMI-III <sup>b</sup>	MCMI-III <sup>a</sup>	MCMI-III <sup>b</sup>
Ansiedad	46	28	57	64	80	75
Somatoforme	24	12	25	24	53	39
Hipomanía	6	4	44	64	53	39
Distimia	38	23	55	65	88	81

Escala	Prevalencia		Sensibilidad		PPP	
	MCMI-III <sup>a</sup>	MCMI-III <sup>b</sup>	MCMI-III <sup>a</sup>	MCMI-III <sup>b</sup>	MCMI-III <sup>a</sup>	MCMI-III <sup>b</sup>
Abuso de alcohol	30	17	65	80	91	88
Abuso de sustancias	19	11	78	82	92	93
Trastorno postraumático	18	6	53	88	73	67
Trastorno del pensamiento	10	4	52	100	59	52
Depresión mayor	20	11	65	84	72	66
Trastorno delirante	9	2	40	50	63	33

*Nota.* PPP = poder predictivo positivo; MCMI-III = Inventario Clínico Multiaxial de Millon; <sup>a</sup> = Datos definitivos de la primera edición del manual del MCMI-III; <sup>b</sup> = Datos recogidos en la segunda edición del Manual del MCMI-III

Llama la atención que en la segunda edición del manual del MCMI-III los datos sobre validez diagnóstica aportados fueran en esta ocasión tan solo los relacionados con la sensibilidad y el poder predictivo positivo, ya que existe suficiente evidencia sobre la necesidad de realizar un análisis más amplio para conocer adecuadamente las propiedades diagnósticas de un instrumento. Sería interesante que los autores hubieran publicado otros estadísticos que proporcionasen información no redundante sobre la eficacia diagnóstica. Hsu (2002) ha calculado a partir de los datos proporcionados por los autores del inventario otros estadísticos como la especificidad, el poder predictivo negativo, el índice kappa y el área bajo la curva ROC (*Receiver Operating Characteristics curve*, o curva de las características operativas del receptor o curva COR). Tales estadísticos, para los dos estudios de las dos ediciones del manual del MCMI-III, se recogen en la Tabla 4.7. Observando los datos de esta tabla, destaca la diferencia entre los índices de ambos estudios. Por ejemplo, en el estudio de la edición de 1997 del manual se encontró, como promedio para todas las escalas del MCMI-III,

una concordancia diagnóstica moderada ( $\kappa = 0.61$ ), mientras que en el caso del estudio de la edición de 1994 del manual tal concordancia media se situaba por debajo de lo aceptable ( $\kappa = 0.128$ ) (véase la Tabla 4.7).

**Tabla 4.7.** Comparación de los estadísticos medios de validez diagnóstica del MCMI-III publicados en sus dos primeros manuales junto con los estadísticos calculados por Hsu (2002) (medias de todas las escalas)

Estadístico de validez diagnóstica	Estudio de 1997	Estudio de 1994
Sensibilidad	0.67	0.275
Especificidad	0.968	0.860
Poder predictivo positivo	0.640	0.223
Poder predictivo negativo	0.964	0.887
Kappa de Cohen	0.61	0.128
Área bajo la curva ROC	0.937	0.653

#### 4.4. Las escalas para el diagnóstico de los trastornos depresivos del MCMI

La presente tesis doctoral se centra en la validez diagnóstica de las escalas de los trastornos depresivos del MCMI-II y del MCMI-III. Las dos escalas del MCMI que valoran estos trastornos son la escala de Depresión Mayor (CC) y la escala de Distimia (D). El contenido y la longitud de ambas escalas han variado a lo largo de las distintas versiones del cuestionario. En el caso de la escala CC, en el MCMI-II tenía 31 ítems y en el MCMI-III se ha reducido a 17 ítems, mientras que la escala D tenía 36 ítems en el MCMI-II y cuenta con 14 ítems en el MCMI-III (véase la Tabla 4.8).

**Tabla 4.8.** Número de ítems de las escalas de Depresión Mayor (CC) y Distimia (D) del MCMI-II y MCMI-III

Versión del MCMI	Depresión Mayor	Distimia
MCMI-II	31	36
MCMI-III	17	14

Las correlaciones entre las dos escalas son muy elevadas, en parte porque tienen un gran solapamiento de ítems, reflejando en este sentido el propio solapamiento de los criterios diagnósticos DSM de los dos trastornos que tratan de evaluar. En el caso de la adaptación española del MCMI-II, se puede observar en la tercera edición del manual de dicha adaptación que la correlación entre la escala D y la CC es de 0.97 (Millon, 2002, adaptación española de A. Ávila-Espada). Por su parte, en el caso de la edición española del MCMI-III (Millon et al., 2007, adaptación española de V. Cardenal y M. P. Sánchez), la correlación entre ambas escalas es de 0.86.

Por otro lado, en la cuarta edición del manual del MCMI-III original se recoge que las correlaciones de las escalas C y DD con el BDI-II son 0.71 y 0.74, respectivamente (Millon et al., 2009).

### **Validez de contenido de las escalas depresivas del MCMI-II y MCMI-III**

En el caso del MCMI-II, su manual señala que en la escala D, “las puntuaciones altas implican que aunque no aparezca en la vida cotidiana del paciente, se ha visto afectado por un período de dos o más años con sentimientos de desánimo o culpabilidad, una carencia de iniciativa y apatía en el comportamiento, baja autoestima y con frecuencia expresiones de inutilidad y comentarios autodesvalorativos. Durante los periodos de depresión, puede haber llantos, ideas suicidas, sentimientos pesimistas hacia el futuro, alejamiento social, apetito escaso o excesivas ganas de comer, agotamiento crónico, pobre concentración, pérdida marca del interés por actividades lúdicas y una disminución de la eficacia en cumplir tareas cotidianas y rutinarias de la vida” (Millon, 1998, p. 38).

Según ese mismo manual, las personas que puntúan alto en la escala CC “son habitualmente incapaces de funcionar en un ambiente normal, se deprimen

gravemente y expresan temor hacia el futuro, ideas suicidas y un sentimiento de resignación. Algunos exhiben un marcado enlentecimiento motor, mientras que otros muestran una cualidad agitada, paseando continuamente y lamentando su estado triste (Millon, 1998, p. 38).

También en el caso del MCMI-III ambas escalas miden sintomatología depresiva, ocurrida últimamente en el caso de la escala CC, o durante los últimos uno o dos años para la escala D (Millon et al., 2007). Los autores señalan de nuevo que “los pacientes con puntuaciones altas en la escala de Distimia siguen involucrados en la vida diaria, pero han estado preocupados durante años con sentimientos de desánimo o culpa, falta de iniciativa apatía conductual y baja autoestima”, mientras que los pacientes con puntuaciones altas en la escala de Depresión Mayor “normalmente son incapaces de funcionar en un ambiente normal, están severamente deprimidos” (Millon et al., 2007, pp. 46 y 47).

Los autores del MCMI han señalado en las sucesivas versiones del inventario su intención de acercar sus escalas a los criterios diagnósticos del DSM. Sanz, Izquierdo y García-Vera (2011) han analizado desde la perspectiva de la validez de contenido los ítems de las escalas CC y D de las dos últimas versiones del MCMI con respecto a los criterios diagnósticos del DSM-IV, encontrando que este objetivo se ha conseguido razonablemente, aunque no por completo (véase la Tabla 4.9).

Concretamente, respecto al MCMI-III, parece que todos los ítems de la escala CC evalúan sintomatología depresiva, pero dichos ítems solamente cubren el 66.7% de los criterios del DSM-IV para el episodio depresivo mayor, por lo que se puede afirmar que la escala CC del MCMI-III no cubre totalmente el contenido del constructo de depresión mayor que propone el DSM-IV. Así, hay 3 criterios del trastorno depresivo mayor en el DSM-IV (agitación o enlentecimiento psicomotor,

sentimientos excesivos o inapropiados de inutilidad o culpa y disminución de la capacidad para pensar, concentrarse o tomar decisiones) que no están recogidos por los ítems de la escala CC del MCMI-III (véanse las Tablas 4.9 y 4.10), aunque en el caso de los sentimientos excesivos o inapropiados de inutilidad o culpa (criterio 7), se puede señalar que en ocasiones la culpa está relacionada con la percepción de estar siendo castigado, por lo que el ítem sobre la pérdida de placer por no merecerlo (ítem 104) de la escala CC del MCMI-III podría de manera secundaria evaluar también esos sentimientos y, por tanto, serían dos los criterios del episodio depresivo mayor del DSM-IV no recogidos por la escala CC del MCMI-III. Por otro lado, existen cuatro criterios del trastorno distímico del DSM-IV que no aparecen reflejados por los ítems de la escala D del MCM-III: aumento/disminución del apetito, insomnio o hipersomnia, dificultades para concentrarse o para tomar decisiones, y sentimientos de desesperanza (Sanz et al., 2011; véase la Tabla 4.9).

En el caso del MCMI-II, el porcentaje de síntomas del episodio depresivo mayor que cubre la escala CC es ligeramente mayor, con un 77.8% de los criterios sintomáticos del DSM-IV cubiertos, pero, no obstante, no se encuentran en esta escala elementos relacionados con cambios en el peso o el apetito (criterio 4) y con los criterios 6 (agitación/enlentecimiento psicomotor) y 8 (disminución de la capacidad para pensar, concentrarse o tomar decisiones). Además, hay un ítem de la escala CC del MCMI-II relacionado con los sentimientos de desesperanza (ítem 150) que tiene concordancia con un criterio sintomático del trastorno distímico del DSM-IV (véase el criterio 11 de la Tabla 4.9), pero que no está recogido por el DSM-IV como síntoma de la depresión mayor. De hecho, el porcentaje de ítems que valoran síntomas depresivos es del 67.7% en esta escala (CC) en el MCMI-II. Es importante recordar que, aunque las escalas del MCMI-II fueron diseñadas con el DSM-III-R como criterio de referencia, los criterios sintomáticos del episodio depresivo mayor son los mismos en el DSM-III-R y en el DSM-IV.



Otra cuestión importante que refleja el análisis de la validez de contenido resumidos en las Tablas 4.9 y 4.10, es que la escala D del MCMI-III cubre un mayor porcentaje de síntomas referidos al diagnóstico de depresión mayor (55.5%) que al diagnóstico de distimia (42.8%). De hecho, la escala de Distimia (D) del MCMI-II es, de todas las escalas de cualquiera de las dos versiones del MCMI (MCMI-II y MCMI-III), la que más porcentaje de los criterios diagnóstico del trastorno depresivo mayor cubre, con un 88.9%, siendo este porcentaje mayor que el de criterios sintomáticos de distimia del DSM-IV cubiertos por dicha escala (71.4% de los criterios).

Por otro lado, en relación con el criterio sintomático del estado de ánimo deprimido (o irritable en el caso de niños y adolescentes) de los trastornos depresivos (criterio 1 de la Tabla 4.9), el número de ítems relacionados con dicho criterio son 6 y 5 en las escalas CC del MCMI-II y del MCMI-III, respectivamente, mientras que en el caso de la escala D del MCMI-II son 10 ítems y en la del MCMI-III son 2 ítems, todo lo cual sugiere que el peso específico que se da a un criterio diagnóstico en relación con los demás es muy diferente en las distintas escalas del MCMI, lo que también influye en su validez de contenido. Del mismo modo, en la escala D del MCMI-II tenía un peso muy significativo el criterio diagnóstico relacionado con los sentimientos de culpa (6 ítems, lo que supone el 16,7% de los ítems de la escala D del MCMI-II), y dicho peso ha sido ligeramente aumentado en el MCMI-III (3 ítems, lo que supone el 21,4% de los ítems de la escala D del MCMI-III).

**Tabla 4.9** Correspondencia entre los síntomas depresivos clave del diagnóstico de los trastornos depresivos y los ítems de las escalas de trastornos depresivos del MCMI-II y del MCMI-III (adaptado de Sanz, Izquierdo y García-Vera, 2011)

Síntomas depresivos de los criterios diagnósticos		Ítems de los cuestionarios de depresión			
DSM-IV (1994) †	CIE-10 (1993) ‡	MCMI-II/Depresión	MCMI-II/Distimia	MCMI-III/Depresión Mayor	MCMI-III/Trastorno Distímico
1. Estado de ánimo deprimido (o irritable en niños y adolescentes)	1. Humor depresivo	(5) Llanto (26) Explosiones de llanto o cólera (36) Llanto (45) Estado de ánimo deprimido (76) Tristeza (99) Desánimo y tristeza	(5) Llanto (26) Explosiones de llanto o cólera (27) Sentirse vacío y solo (36) Llanto (45) Estado de ánimo deprimido (76) Tristeza (99) Desánimo y tristeza (107) Irritabilidad (132) Tristeza y melancolía (167) Sentirse destrozado	(44) Estado de ánimo deprimido y triste (34) Perder los nervios (128) Sentimientos de depresión sin razón (150) Sentimientos de depresión por pensar en el futuro (142) Sentirse vacío y hueco	(62) Tristeza y desánimo (86) Estado de ánimo triste, deprimido y sin ánimo
2. Disminución del placer o del interés	2. Pérdida de interés o placer	(81) Desinterés por el sexo	83) Desinterés social (86) Capacidad de interesarse	(104) Pérdida de placer por no merecerlo (111) Pérdida de interés (148) Pérdida de placer	(104) Pérdida de placer por no merecerlo (111) Pérdida de interés (148) Pérdida de placer
3. Fatiga o pérdida de energía	3. Falta de vitalidad o aumento de fatigabilidad	(33) Fatiga (53) Pérdida de energía (96) Fatiga	(71) Fatiga (53) Pérdida de energía	(1) Quedarse sin fuerzas (4) Sentirse débil y cansado	(55) Sentirse agotado (130) Pérdida de energía para

Síntomas depresivos de los criterios diagnósticos		Ítems de los cuestionarios de depresión			
		(108) Pérdida energía	(96) Fatiga (108) Pérdida energía	(55) Sentirse agotado (130) Pérdida de energía para concentrarse en responsabilidades diarias (107) Pérdida de apetito y dificultad para dormir	concentrarse en responsabilidades diarias
4. Aumento/disminución del peso o del apetito	4. Aumento/disminución del apetito y peso				
5. Insomnio o hipersomnia	5. Alteraciones del sueño	(72) Insomnio	(72) Insomnio	(74) Dificultad para dormir y sueño no reparador (107) Pérdida de apetito y dificultad para dormir (149) Temblores y dificultad para dormir por recuerdo doloroso	
6. Agitación o enlentecimiento psicomotor	6. Cambios de actividad psicomotriz con agitación o inhibición		*(46) Perfeccionismo y enlentecimiento		
7. Sentimientos excesivos o inapropiados de inutilidad o culpa	7. Auto-reproches desproporcionados y sentimientos de culpa excesiva e inadecuada	(54) Sentimientos de fracaso (95) Impulsividad y culpa (110) Sentimientos de culpa (136) Culpa e ideas de suicidio	(54) Sentimientos de fracaso (97) Sentimientos de culpa (110) Sentimientos de culpa (136) Culpa e ideas		(24) Sentimientos de fracaso (25) Sentimientos de culpa (56) Culpabilidad por inutilidad

Síntomas depresivos de los criterios diagnósticos		Ítems de los cuestionarios de depresión		
		(154) Sentimientos de castigo	de suicidio (154) Sentimientos de castigo (166) Sentimientos de competencia	
8. Disminución de la capacidad para pensar, concentrarse o tomar decisiones	8. Disminución de la capacidad de pensar y concentrarse, acompañadas de indecisión y vacilación	(109) Necesidad de pensar las cosas varias veces sin motivo	(109) Necesidad de pensar las cosas varias veces sin motivo	
9. Ideas recurrentes de muerte o suicidio o intentos suicidas	9. Pensamientos de muerte o suicidio, o conducta suicida	(59) Pensamientos de suicidio (79) Pensamientos de suicidio (136) Culpa e ideas de suicidio	(59) Pensamientos de suicidio (79) Pensamientos de suicidio (136) Culpa e ideas de suicidio	(154) Intentos suicidas (171) Ideación suicida
10. Baja autoestima	10. Pérdida de confianza y autoestima, y sentimientos de inferioridad		(25) Inseguridad sobre lo que es y quiere (166) Sentimientos de competencia	(141) Sentimientos de que los demás tienen mala opinión de uno mismo
11. Sentimientos de desesperanza	11. Sentimientos de desesperanza o desesperación			(150) Sentimientos de depresión por pensar en el futuro
	12. Sentimiento de			(130) Pérdida de energía (130) Pérdida de

Síntomas depresivos de los criterios diagnósticos		Ítems de los cuestionarios de depresión	
incapacidad para enfrentarse con las responsabilidades de la vida diaria		para concentrarse en responsabilidades diarias	energía para concentrarse en responsabilidades diarias
13. Llanto fácil	(5) Llanto (26) Explosiones de llanto o cólera (36) Llanto Criterio Ausente en Depresión Mayor (DSM-IV)	(5) Llanto (26) Explosiones de llanto o cólera (36) Llanto Criterio ausente en Distimia (DSM-IV)	
14. Pesimismo sobre el futuro o cavilaciones sobre el pasado		Criterio ausente en Distimia (DSM-IV)	(15) Pesimismo
15. Aislamiento social		(27) Sentirse vacío y solo	
16. Disminución de la locuacidad			

*Nota.* Entre paréntesis los números de los correspondientes ítems en cada versión del inventario. † Del 1 al 9 aparecen los criterios sintomáticos del episodio depresivo mayor, mientras que los del trastorno distímico corresponden a los números 1, 3, 4, 5, 8, 10 y 11. ‡ Del 1 al 10 aparecen los criterios sintomáticos del episodio depresivo, mientras que los de la distimia corresponden a los números 1, 2, 3, 5, 8, 10, 11, 12, 13, 14, 15 y 16.

**Tabla 4.10.** Porcentajes de síntomas depresivos cubiertos por las escalas de trastornos depresivos del MCMI-II y del MCMI-III y porcentajes de ítems de dichas escalas que evalúan síntomas depresivos (adaptado de Sanz, Izquierdo y García-Vera, 2011)

Inventario/ Escala	Criterios diagnósticos del DSM-IV			Criterios diagnósticos de la CIE-10			% de ítems que evalúan síntomas depresivos
	% de síntomas de depresión mayor	% de síntomas de distimia	% de síntomas atípicos o de especificaciones	% de síntomas de depresión mayor	% de síntomas de distimia	% de síntomas atípicos o de especificaciones	
MCMI-II/CC	77,8 %	57,1%	7,1%	70%	50%	0%	67,7%
MCMI-II/D	88,9 %	71,4%	7,1%	90%	66,7%	0%	83,3%
MCMI- III/CC	66,7%	71,4%	0%	60%	50%	0%	100%
MCMI-III/D	55,5%	42,8%	0%	50%	50%	0%	94,1%

*Nota.* CC = Escala de Depresión Mayor; D = Escala de Distimia.

En conclusión, aunque Millon y sus colaboradores han declarado que en el MCMI-II y el MCMI-III existe un esfuerzo por acercar el contenido de las escalas clínicas a los criterios diagnósticos del DSM, la sintomatología recogida por las escalas de Depresión y Distimia de ambos inventarios no se corresponde totalmente con los criterios sintomáticos del DSM-IV para los respectivos trastornos depresivos (y, por tanto, tampoco recogerá totalmente los criterios sintomáticos previstos en el desarrollo del DSM-5 y que parece que no cambiarán respecto a los del DSM-III-R y DSM-V). Además, tomando como criterio los criterios sintomáticos del DSM-IV, el contenido de las escalas de Depresión Mayor y Distimia no es totalmente específico al trastorno en cuestión, de manera que la escala de Distimia de ambos inventarios cubre más criterios diagnósticos del episodio depresivo mayor que del trastorno distímico, mientras que la escala de Depresión Mayor del MCMI-III cubre más criterios diagnósticos del trastorno distímico que del episodio depresivo mayor. Por último, en comparación al MCMI-II, el MCMI-III no supone una mejora significativa respecto a la validez de contenido de las escalas de Depresión Mayor y Distimia en lo referido a la cobertura de síntomas diagnósticos de los correspondientes trastornos del DSM-IV. Sin embargo, el MCMI-III sí supone una mejora significativa respecto al MCMI-II en lo tocante al porcentaje de ítems de ambas escalas que evalúan síntomas depresivos. No obstante, a pesar de las limitaciones señaladas, se puede afirmar que en general las escalas de Depresión Mayor y Distimia del MCMI-II y del MCMI-III cubren razonablemente bien los síntomas que conforman las definiciones de episodio depresivo mayor y de trastorno distímico del DSM-IV, las cuales se erigen actualmente, junto con las definiciones de la CIE-10, como las definiciones más validas y consensuadas de los trastornos depresivos.

#### 4.5. La validez diagnóstica de las escalas para el diagnóstico de los trastornos depresivos del MCMI

Como se ha señalado, el MCMI-I dejó muchas dudas respecto a su utilidad como instrumento diagnóstico, especialmente la escala de Depresión Psicótica. Los análisis de la validez diagnóstica del MCMI-II recogidos en su manual original (Millon, 2002) señalaban una mejoría en dicha escala, que pasó a llamarse de Depresión Mayor después de unas modificaciones significativas (véase la Tabla 4.11).

**Tabla 4.11.** Índices de eficacia diagnóstica de las escalas de Depresión Mayor y Distimia del MCMI-II (puntos de corte de TB  $\geq 75$  y de TB  $\geq 85$ ) para los correspondientes trastornos depresivos del DSM-III-R (Millon, 2002)

Escala (punto de corte)	Prevalencia	Sens.	Espe.	PPP	PPN	PDT
Distimia (TB 75)	46	81	83	80	84	82
Depresión mayor (TB 85)	14	68	96	73	95	92
Distimia (TB 75)	29	76	88	72	90	84
Depresión mayor (TB 85)	6	47	97	50	97	94

*Nota.* Sens = Sensibilidad; Espe = Especificidad; PPP = poder predictivo positivo; PPN = poder predictivo negativo; PDT = poder diagnóstico total.

Como puede observarse en la Tabla 4.11, en el caso de la escala D todos los valores comentados sobre la eficacia diagnóstica fueron superiores a 0.72, mientras que de nuevo la escala de CC presentó unos datos inferiores. En el caso del punto de corte correspondiente a TB 85 se encontró para la escala CC una sensibilidad y poder predictivo positivo de 0.47 y 0.50, respectivamente. Tres aspectos que llaman la atención en ese estudio respecto al análisis de la eficacia diagnóstica de las escalas de trastornos depresivos del MCMI-II son los siguientes:

1. La escala D del MCMI-II se analiza con unos datos de prevalencia del trastorno distímico del 46%, 3 veces superior a la prevalencia de la



depresión mayor y superior al conjunto de los trastornos de ansiedad. De hecho, es el trastorno más prominente en el 29% de los casos, casi 5 veces más que el trastorno depresivo mayor. Estos datos no concuerdan con los estudios epidemiológicos realizados en población general (Alonso et al., 2010) ni con otros estudios sobre el MCMI (Saulsman, 2011; Rossi y Sloore, 2005).

2. Los puntos de corte seleccionados en este caso podrían no ser los que mayor eficacia diagnóstica proporcionan. Por ejemplo, en la escala de Depresión Mayor, en los sujetos con puntuaciones  $TB \geq 75$ , la especificidad y el poder predictivo negativo son muy altos (96% y 95% respectivamente), por lo que aumentar el punto de corte hasta la puntuación correspondiente a la TB 85 no añade prácticamente más especificidad (97%) ni poder predictivo negativo (97%). Sin embargo, la sensibilidad disminuye muy significativamente (desde el 67% en el caso de la TB 75 al 47% en la TB 85), así como el poder predictivo negativo (desde el 73% al 50%). Asumiendo que, como se ha comentado, los autores dan prioridad a la especificidad sobre la sensibilidad, quizá un análisis detallado de la eficacia diagnóstica podría asignar puntos de corte más equilibrados en esta escala.
3. En el manual no se publicaron otros estadísticos o índices de validez diagnóstica tan importantes como la concordancia (kappa) o el área bajo la curva ROC.

En una publicación posterior se analizó la capacidad de las escalas CC y D del MCMI-II para detectar el trastorno depresivo mayor, en una muestra psiquiátrica de 109 sujetos diagnosticados con un trastorno depresivo (Piersma, 1991). Los datos aportados por los autores muestran una mejoría con respecto al MCMI-I: insuficiente en la escala CC y peores cualidades psicométricas con respecto a la

escala D en el diagnóstico del trastorno depresivo mayor (véase la Tabla 4.12). En este caso los datos sobre especificidad son menores de lo esperado.

**Tabla 4.12.** Estadísticos de validez diagnóstica de las escalas de Depresión Mayor y Distimia del MCMI-II para diagnosticar trastorno depresivo mayor en el estudio de Piersma (1991)

Escala	Sensibilidad	Especificidad	PPP	PPN	PDT	Prev.
Depresión Mayor (CC)	61	52	81	28	59	77
Distimia (D)	86	32	81	40	73	77

*Nota.* PPP = poder predictivo positivo; PPN = poder predictivo negativo; PDT = poder diagnóstico total; Prev. = prevalencia.

En otro estudio sobre esta versión del inventario, se aplicó el MCMI-II a 110 personas y el MCMI-I a 148, dentro de una muestra psiquiátrica. Se realizaron diagnósticos independientes del resultado del inventario por parte de clínicos expertos, agrupando en una sola categoría los diagnósticos de los trastornos depresivos (Wetzler y Marlowe, 1993).

**Tabla 4.13.** Índices de validez diagnóstica de las escalas de Depresión Psicótica o Mayor y de Distimia del MCMI-I y del MCMI-II para diagnosticar un trastorno depresivo en el estudio de Wetzler y Marlowe (1993)

Escala y punto de corte	Sens.	Espe.	PPP	PPN	PDT	Prev.
MCMI-I Distimia $\geq$ TB 75	89	53	83	65	71	49
MCMI-I Distimia $\geq$ TB 85	74	73	74	73	74	49
MCMI-I Depresión Psicótica $\geq$ TB 75	16	87	52	55	52	49
MCMI-I Depresión Psicótica $\geq$ TB 85	08	93	51	55	51	49
MCMI-II Distimia $\geq$ TB 75	57	70	57	69	63	55
MCMI-II Distimia TB $\geq$ 85	33	86	52	74	57	55
MCMI-II Depresión Mayor $\geq$ TB 75	18	86	47	61	49	55
MCMI-II Depresión Mayor $\geq$ 85	10	92	46	60	47	55

*Nota.* Sens = Sensibilidad; Espe = Especificidad; PPP = poder predictivo positivo; PPN = poder predictivo negativo; PDT = poder diagnóstico total; Prev. = prevalencia.

Los datos de este estudio, que se recogen en la Tabla 4.13, son acordes con las publicaciones anteriores en cuanto a que se aprecia una cierta mejoría en la escala CC del MCMI-II con respecto al MCMI-I en especificidad (0.86 en el caso del  $TB \geq 75$ ), aunque los datos sobre sensibilidad continúan siendo excesivamente bajos para considerarla un instrumento válido para el diagnóstico de los trastornos depresivos (0.18 en el caso de la  $TB \geq 75$ ). En el caso de la escala D, parece que muestra de nuevo mejores propiedades diagnósticas que la escala CC, con una sensibilidad mayor a la prevalencia del trastorno (0.57) en el caso del punto de corte  $TB \geq 75$  (véase la Tabla 4.13).

### MCMI-III

Como ya se ha comentado, se han publicado los resultados sobre la eficacia diagnóstica de las escalas de los trastornos depresivos del MCMI-III en las dos primeras versiones de su manual original. En el primero de ellos, los datos proporcionados por Millon señalaban datos de eficacia diagnóstica bajos (véase la Tabla 4.14). Los autores del manual señalaban que se trataban de datos preliminares, por la que los datos definitivos fueron aportados en el manual publicado en 1997 (véase la Tabla 4.14).

**Tabla 4.14.** Prevalencia, sensibilidad y poder predictivo positivo (PPP) de las escalas de Distimia y Depresión Mayor del MCMI-III con el punto de corte de  $TB \geq 85$  en el estudio de Millon, 1997)

	Prevalencia		Sensibilidad		PPP	
	MCMI-III <sup>a</sup>	MCMI-III <sup>b</sup>	MCMI-III <sup>a</sup>	MCMI-III <sup>b</sup>	MCMI-III <sup>a</sup>	MCMI-III <sup>b</sup>
Distimia	38	23	55	65	88	81
Depresión mayor	20	11	65	84	72	66

*Nota.* PPP = Poder predictivo positivo; <sup>a</sup> = Datos definitivos de la primera edición del manual del MCMI-III; <sup>b</sup> = Datos recogidos en la segunda edición del Manual del MCMI-III

Ante la ausencia en las dos primeras ediciones de los manuales de algunos índices significativos de eficacia diagnóstica de las escalas del MCMI-III, otros autores se han encargado de calcularlos (Rezlaiff, 1996; Hsu, 2002). Los cálculos se realizaron sobre el punto de corte de TB 85 y se recogen en la Tabla 4.15 y en la Tabla 4.16.

**Tabla 4.15.** Estadísticos de validez diagnóstica de las escalas de Depresión Mayor y Distimia del MCMI-III calculados por Rezlaiff (1996) a partir de los datos publicados en el primer manual del MCMI-III (Millon, 1994)

Escola	Prev.	Sensibilidad	Especificidad	PPP	PPN
Depresión Mayor	21.5	56,6	73.0	36	86
Distimia	36.2	36.7	58.8	34	62

*Nota.* Prev. = prevalencia; PPP = poder predictivo positivo; PPN = poder predictivo negativo.

**Tabla 4.16.** Estadísticos sobre la validez diagnóstica de las escalas de Depresión Mayor y Distimia del MCMI-III calculados por Hsu (2002) a partir de los datos de las dos versiones del manual del MCMI-III (Millon, 1994, 1997)

Escola (año de edición del manual)	PPP	PPN	Kappa	Área bajo la curva ROC
Depresión Mayor (1994)	36		0.24	0.71
Depresión Mayor (1997)	66	98	0.70	0.97
Distimia (1994)	33		-0.04	0.46
Distimia (1997)	081	90	0.65	0.90

*Nota.* PPP = poder predictivo positivo; PPN = poder predictivo negativo.

Analizando los datos de las Tablas 4.15 y 4.16, se puede observar una mejoría en los datos de validez diagnóstica de las escalas CC y D del MCMI-III, sobretudo en el caso de la escala de Depresión Mayor. Los datos analizados por

Hsu (2002) y Rezlaff (1996) son los que el propio Millon consideró preliminares en la versión del manual de 1994.

### Otras investigaciones con la versión original del MCMI-III sobre la validez diagnóstica de las escalas de Distimia y Depresión Mayor

En una investigación sobre las escalas de trastornos clínicos del MCMI-III, Gibeau y Choca (2005) estudiaron sus índices de validez diagnóstica teniendo en cuenta los baremos de corrección de la segunda edición del manual del MCMI-III (véase la Tabla 4.17). En el estudio participó una muestra de más de 371 pacientes de los que la práctica totalidad eran hombres (355), puesto que el estudio se realizó en población de veteranos de guerra. Los diagnósticos fueron realizados sin que los especialistas en salud mental tuvieran conocimiento de las puntuaciones del MCMI-III. El estudio aportó datos sobre la validez diagnóstica de las escalas de Depresión Mayor y Distimia del MCMI-III que muestran en general un poder predictivo aceptable y una baja sensibilidad (véase la Tabla 4.17).

**Tabla 4.17.** Estadísticos de validez diagnóstica de las escalas de Depresión Mayor y Distimia del MCMI-III en el estudio de Gibeau y Choca (2005)

Escala y punto de corte	Prevalencia	Sensibilidad	Especificidad	PPP	PPN
Depresión Mayor TB $\geq$ 75	42	45	79	60	67
Depresión mayor TB $\geq$ 85	42	26	90	67	63
Distimia TB $\geq$ 75	42	73	58	56	75
Distimia TB $\geq$ 85	42	32	86	62	63

*Nota.* PPP = poder predictivo positivo; PPN = poder predictivo negativo.

Es destacable que en el estudio de Gibeau y Choca (2005) se compararon los resultados de las escalas CC y D del MCMI-III con los diagnósticos de los trastornos depresivos en su conjunto, no con el trastorno depresivo mayor y el trastorno distímico de manera aislada. En línea con lo encontrado en los diferentes estudios de Millon comentados con anterioridad, los datos de especificidad fueron mayores que los de sensibilidad (Gibeau y Choca, 2005; véase la Tabla 4.16). Dependiendo del uso que tenga el cuestionario, se puede considerar que siguiendo este estudio sería más interesante utilizar las puntuaciones  $TB \geq 75$  en la medida en que ofrecen estadísticos más equilibrados (tienen buena sensibilidad sin perder especificidad). Es más, la sensibilidad de las escalas CC y D del MCMI-III con  $TB \geq 85$  es menor que la prevalencia del trastorno, lo que indica una baja validez diagnóstica. Hsu (2002) comenta que en estos casos los instrumentos tienen peor validez diagnóstica que tirar una moneda al aire para realizar el diagnóstico. Craig (2005a) señala que esto no es cierto, puesto que tirar una moneda supone diagnosticar a un 50% de las personas con un mismo trastorno, lo que claramente supone cometer más errores (puesto que ningún trastorno tiene una prevalencia mayor del 50%). Lo que es cierto, no obstante, es que un dato de sensibilidad inferior a su prevalencia supone un mal dato de validez diagnóstica.

Curiosamente, Gibeau y Choca (2005) no publicaron otros estadísticos de validez diagnóstica también importantes como, por ejemplo, el índice kappa o el área bajo la curva ROC, incluso después de reflejar en su propio artículo las críticas que recibieron al respecto las publicaciones de Millon.

Por último, recientemente se llevó a cabo un estudio para evaluar la validez de constructo y la eficacia diagnóstica de las escalas de Depresión Mayor, Distimia y Ansiedad del MCMI-III en Australia (Saulsman, 2011). Sobre una población de usuarios de los servicios de salud mental en ese país, se seleccionó una muestra

de pacientes con problemas de ansiedad y estado de ánimo. Se eliminaron del estudio aquellos pacientes con sintomatología psicótica y con abuso o dependencia de alcohol y otras sustancias. Se analizaron las historias de 696 pacientes de manera retrospectiva, encontrando que más de la mitad de los sujetos tenían más de un trastorno. Los diagnósticos se realizaron durante la primera consulta mediante la entrevista MINI, sin que los evaluadores conocieran las puntuaciones del MCMI-III, ni de ningún otro cuestionario.

La correlación entre el BDI-II y las escalas D y CC fueron elevadas, 0.67 y 0.69, respectivamente, indicando una buena validez convergente. Respecto al BAI, las correlaciones fueron menores, 0.25 en el caso de la escala D y 0.31 en la escala CC, señalando una buena validez discriminante (Saulsman, 2011).

Los análisis sobre la eficacia diagnóstica fueron realizados con los datos de los pacientes para los que se le había asignado un solo diagnóstico, sin comorbilidad. Por tanto, los datos se analizaron sobre una submuestra de 284 sujetos con dos grupos diagnósticos: 156 pacientes con diagnóstico de trastorno depresivo (129 con diagnóstico de depresión mayor y 26 de distimia) frente a 119 pacientes con diagnóstico de trastorno de ansiedad. Los resultados de esos análisis se resumen en la Tabla 4.18.

**Tabla 4.18.** Estadísticos de validez diagnóstica de las escalas de Depresión Mayor y Distimia del MCMI-III en el estudio de Saulsman (2011)

Escala y punto de corte	Sens.	Espe.	PPP	PPN	AUC
Distimia TB $\geq$ 75	66	64	72	58	0.65
Distimia TB $\geq$ 85	21	95	85	46	0.58
Depresión Mayor TB $\geq$ 75	55	87	85	58	0.71
Depresión Mayor TB $\geq$ 85	35	91	84	50	0.63

*Nota.* Sens. = Sensibilidad; Espe. = Especificidad; PPP = poder predictivo positivo; PPN = poder predictivo negativo; AUC = Área bajo la curva ROC.

Como se puede observar en la Tabla 4.18, cuando se trata de valorar el conjunto de los trastornos depresivos frente al diagnóstico de trastorno de ansiedad, la escala de Depresión Mayor tenía un área bajo la curva ROC de 0.71, con datos aceptables de sensibilidad de 0.55 y poder predictivo negativo de 0.57. Según la autora, “la escala de Distimia..., mostró pobres características operativas para diferenciar pacientes con ansiedad y depresión. La escala de Depresión Mayor mostró mejores resultados con la puntuación de corte de TB 75; no obstante, aun así esta escala mostraba una eficiencia diagnóstica solamente pasable, con buena especificidad y poder predictivo positivo, pero con una sensibilidad y poder predictivo negativo menos que óptimos” (Saulsman, 2011, pp. 80-81).

Es destacable que cuando los análisis se realizaron para los diagnósticos separados de distimia y trastorno depresivo mayor, la autora encontró valores semejantes de eficacia diagnóstica a los anteriores. En el caso de la escala de Depresión Mayor, la sensibilidad apenas varió tanto para la TB  $\geq 75$  (0.56) como para TB  $\geq 85$  (0.36). Por otro lado, la escala de Distimia sí disminuyó en sus cualidades de eficacia diagnóstica, con una sensibilidad de 0.64 para TB  $\geq 75$  y de solamente 0.19 para la TB  $\geq 85$ . En este estudio, el 61.2% de los pacientes puntuaba en la escala de Distimia por encima de la TB 75, frente a un 42.2% en el caso de la escala de Depresión Mayor, mientras que la prevalencia del trastorno depresivo mayor según la entrevista MINI fue muy similar, con un 42.7%, pero no en el caso de la distimia, puesto que la MINI detectó sólo una prevalencia de 11.3% (Saulsman, 2011).

Saulsman (2011, p. 81) afirma que la escala de Depresión Mayor mostró en este estudio un mejor comportamiento que los señalados en las versiones anteriores del MCMI, siendo en la TB  $\geq 75$  la única escala de las tres evaluadas que



“mostró una habilidad aceptable para distinguir adecuadamente los grupos diagnósticos”. Considera que al excluir los pacientes con comorbilidad, “las condiciones de este estudio favorecieron la sobreestimación de la eficiencia diagnóstica de las escalas de Ansiedad, Distimia y Depresión Mayor del MCMI-III, mostrando lo que se puede conseguir en condiciones diagnósticas sencillas. Incluso en estas condiciones favorables, el MCMI-III no tuvo un buen rendimiento general” (Saulsman, 2011, p. 81).

En cualquier caso, los datos disponibles de validez diagnóstica de la versión original del MCMI-III permiten a los psicólogos que utilizan dicha versión conocer las características de dichas escalas como instrumentos diagnósticos y valorar sus puntos fuertes y débiles al respecto con una base metodológica y en función de unos índices empíricos de validez diagnóstica concretos, lo cual no siempre es así respecto a las adaptaciones del MCMI en otros idiomas.

### **Adaptaciones del MCMI a otros idiomas**

Para la adaptación del MCMI a otros idiomas, las propiedades estadísticas de la versión original del MCMI no se pueden asumir directamente, sino que han de determinarse de manera empírica. Además de las consideraciones generales sobre la adaptación de los test, en este inventario en particular es especialmente importante tomar en consideración la prevalencia de los trastornos, ya que las puntuaciones están transformadas en TB. En cada país, esta prevalencia es distinta (Alonso et al., 2004a; Haro et al., 2006), por lo que se debe calcular de nuevo en cada población y a partir de ella elaborar los baremos y analizar las propiedades psicométricas del MCMI.

Hasta 2008 existían al menos 8 traducciones del MCMI-III a otros idiomas (Weiner y Craighead, 2010). Existen en Europa 3 adaptaciones del MCMI-III que cumplen los criterios necesarios de traducción (danés, holandés y belga), además

de la española (Millon et al., 2007, adaptación española de V. Cardenal y M. P. Sánchez), pero, entre las versiones en idiomas distintos del inglés, solamente se han realizado estudios sobre la validez diagnóstica del MCMI para la traducción belga. De hecho, en esta última adaptación, se ha realizado un esfuerzo por superar las limitaciones de las investigaciones sobre la validez diagnóstica del MCMI-III realizadas por Millon (Rossi y Sloore, 2005). Para ello, se llevó a cabo un estudio con una muestra de 627 pacientes de población forense, a los que se les asignaron los diagnósticos tanto del Eje I como del Eje II. Los evaluadores, que no conocían las puntuaciones del MCMI-III de los sujetos, disponían de un listado de criterios detallados sobre los diagnósticos DSM-IV, las descripciones de las escalas del MCMI-III que aparecen en los manuales y las descripciones prototípicas propias de la teoría de Millon (Rossi y Sloore, 2005). Al igual que en la versión original (Millon et al., 1997), los evaluadores asignaron a los diagnósticos una puntuación de 1 a 9 indicando “el grado en el que el síntoma clínico está presente” (1 = ligero, 3 = leve, 5 = presencia de síndrome, 7 = síndrome marcado, 9 = síndrome extremo). “Una puntuación  $\geq$  a 5 implica que el síntoma es suficientemente problemático para justificar un diagnóstico clínico” (Rossi y Sloore, 2005, p. 198). Para completar el proceso diagnóstico en este estudio los evaluadores utilizaron al menos 4 horas de consulta con el sujeto y se utilizó la metodología LEAD (longitudinal expert evaluation that uses all data) o evaluación experta longitudinal que utiliza todos los datos disponibles (Spitzer, 1983). Los autores del estudio encontraron, tras realizar un análisis de las curvas ROC, que existían puntuaciones que ofrecían mejor rendimiento diagnóstico que las puntuaciones TB. De hecho, la adaptación belga detectó la necesidad de realizar estudios de validez diagnóstica para cada adaptación del MCMI a partir de las incongruencias encontradas en las puntuaciones TB de los estudios sobre validez de la adaptación (difícilmente explicables por diferencias culturales), de manera que la adaptación proporcione a

los clínicos el máximo de la información disponible para la interpretación del inventario.

Mientras que Millon (1994, 1997) destaca la importancia del estadístico poder predictivo positivo a la hora de valorar el rendimiento diagnóstico de un instrumento ya que permitiría cometer el mínimo número posible de errores positivos, en la adaptación belga se considera igual de importante disponer del resto de estadísticos relacionados con la validez diagnóstica para así poder hacer una valoración adecuada del rendimiento diagnóstico del instrumento.

En el estudio belga se realizó el análisis de eficacia diagnóstica a partir de la diferenciación que los clínicos hacían entre presencia del trastorno (puntuaciones  $\geq 5$ ) y presencia de síntomas de un trastorno (puntuaciones entre 1 y 9). En el primer caso se comparaba con los sujetos con puntuaciones TB  $\geq 85$  y en el segundo con puntuaciones TB  $\geq 75$ .

**Tabla 4.19.** Estadísticos de validez diagnóstica de la escala de depresión mayor de la adaptación belga del MCMI-III para los diagnósticos asignados con puntuaciones  $\geq 1$  (trastorno ligero) respecto TB  $\geq 75$  o  $\geq 5$  (presencia de síndrome) respecto TB  $\geq 85$  (Rossi y Sloore, 2005)

Tasa Base	Prev.	Sens.	Espe.	PPP	PPN	Kappa	Área ROC
Distimia TB $\geq 75$	42	70	47	49	68	0.16	0.645
Depresión Mayor TB $\geq 75$	47	73	51	57	68	0.243	0.684
Distimia TB $\geq 85$	25	66	59	36	83	0.205	0.678
Depresión Mayor TB $\geq 85$	29	73	55	40	83	0.23	0.687

*Nota.* Prev = Prevalencia; Sens = Sensibilidad; Espe = Especificidad; PPP = Poder predictivo positivo; PPN = Poder predictivo negativo; Kappa = Índice kappa; Área ROC = Área bajo la curva ROC.

Al analizar los resultados obtenidos en la adaptación belga para las escalas D y CC (véase la Tabla 4.19), se encontraron buenos datos de sensibilidad y aceptables datos de especificidad, mientras que los datos sobre el poder predictivo negativo fueron superiores a los relacionados con el poder predictivo positivo. Los valores del índice kappa fueron bajos y el área de la curva ROC se situó significativamente por encima de 0.5.

Rossi y Sloore (2005) señalan que el MCMI-III, en su adaptación belga, tiene buenos datos de validez diagnóstica general y es un instrumento que funciona bien como *screening*, pues, en función de sus índices de validez, permitiría evitar los falsos negativos, pero, para realizar un diagnóstico se recomienda su utilización junto a una recogida de información suplementaria. Las autoras de esta adaptación señalan que la versión original del MCMI-III tiene la intención de priorizar la utilización del test para usos forenses, puesto que la prioridad es evitar falsos positivos, pero, curiosamente, el estudio belga se realizó precisamente con población forense. En cualquier caso, dado que el estudio belga se realizó con ese tipo de población, no se debería asumir directamente que sus resultados son generalizables a otros tipos de población como, por ejemplo, la población clínica que acude habitualmente a los centros de salud mental o los servicios de salud mental de los hospitales generales.

#### **4.6. Problemas de las adaptaciones españolas del MCMI-II y MCMI-III como instrumentos diagnósticos de los trastornos depresivos**

El MCMI es un instrumento que ha sido editado en varias ocasiones en España. En concreto, existen 3 ediciones del MCMI-II (Millon, 1998, 1999 y 2002, adaptación española de A. Ávila-Espada) y una edición del MCMI-III (Millon et al., 2007, adaptación española de V. Cardenal y M. P. Sánchez).

En el caso de la adaptación española del MCMI-II, existen dificultades para la interpretación de las puntuaciones del inventario. En la versión original del instrumento, sus autores tuvieron interés en comprobar la validez diagnóstica de las escalas del MCMI-II con respecto a las puntuaciones TB, pero esto no se hizo en la adaptación española. Además, en las tres versiones españolas del MCMI-II existen discrepancias importantes que hacen difícil saber qué TB se debería tener en cuenta (Sanz, 2007). Las dificultades surgen al encontrar diferencias importantes en las tablas de transformación de las puntuaciones directas en puntuaciones TB en los distintos manuales del MCMI-II y en relación con las tablas del manual original (véase la Tabla 4.20).

**Tabla 4.20.** Correspondencia en cuanto a la definición de las puntuaciones de TB de la escala de Depresión Mayor (CC) en las muestras normativas de mujeres entre la información que aparece en los baremos de los Anexos del manual de la adaptación española del MCM-II y la información que aparece en las tablas de dicho manual

	Versión Original		Versión Española					
	EEUU		1ª y 2ª edición				3ª edición	
	Muestra clínica de mujeres		Muestra clínica de mujeres		Muestra de mujeres con toxicomanía		Muestra clínica de mujeres	
	Baremos	Tablas	Baremos	Tablas	Baremos	Tablas	Baremos	Tablas
TB	PD	PD	PD	PD	PD	PD	PD	PD
85	44	44	50	71	50	n.e.	35	35
75	36	36	42	63	42	n.e.	31-32	31
60	17	17 <sup>a</sup>	23-31	50	23-31	50	24-25	21
35	5	n.e.	11	8	11	8	13-14	n.e.

*Nota.* TB = Puntuaciones tasa base; PD = Puntuaciones directas. n.e. = No se especifican estos valores en el manual de la adaptación española del MCMI-II (1ª ed., 2ª ed. o 3ª ed.). <sup>a</sup> = Media de los pacientes (varones + mujeres) de la muestra de baremación estadounidense.

1. En las dos primeras versiones de los manuales del MCMI-II editadas en España existen discrepancias entre los baremos recogidos en los anexos y las tablas contenidas en el interior del manual (Tablas 5.3 y 5.4 del manual). Estas discrepancias se refieren a los valores de las puntuaciones directas de cada escala que se asignan a las distintas puntuaciones TB. Según apunta Sanz (2007, p. 295):

“Posiblemente, estas discrepancias tienen que ver con errores de tipo tipográfico, pero no queda claro si los errores están en las tablas, en los baremos de los Anexos o en ambos, aunque dados los baremos estadounidenses y los baremos presentados en la 3ª edición del manual, es muy probable que estos errores estén en las tablas”.

Como se puede ver en la Tabla 4.19, en las dos primeras versiones del MCMI-II en España, una puntuación directa de 50 en una mujer se correspondería con una puntuación TB mayor de 85 (presencia de trastorno más prominente) según los baremos de los anexos y sin embargo se correspondería con la TB 60 en las tablas del interior del manual (media de la población clínica).

2. Existe una gran discrepancia entre la correspondencia de las PD y TB en la muestra original estadounidense con respecto a las dos primeras versiones españolas, de manera que es necesario tener una PD mayor en la adaptación española para tener la misma TB. Aunque se puede argumentar que existen razones culturales que justifiquen tales diferencias, el tamaño de las diferencias parece exceder esta posibilidad (Sanz, 2007). En la versión española del MCMI-II se asumió automáticamente la prevalencia de los estudios originales para obtener las TB 75 y TB 85, pero la prevalencia de los trastornos suelen diferir en distintos países (Alonso et al., 2004a; Haro et al., 2006).

3. Los baremos de la 3ª edición del MCMI-II se realizaron sobre una nueva muestra compuesta por 710 sujetos. Como se puede observar, ya no existe prácticamente discrepancia alguna entre los baremos de los anexos y las tablas del interior del manual, pero, sin embargo, la diferencia entre estos datos y los proporcionados por las dos primeras ediciones son muy notables. Por ejemplo, en la escala CC se puede observar que un mujer con una PD de 40 le corresponde una TB mayor de 85 en este último manual, mientras que en las dos primeras versiones, se le asignaría una TB menor de 75. Por tanto, los usuarios del MCMI-II en su versión española deberían aclarar qué manual están utilizando para la corrección del inventario y qué baremo de transformación a TB en concreto.
4. También en el caso de la 3ª adaptación al castellano los baremos de corrección difieren en gran medida de los baremos originales. Podemos ver que el tamaño de estas diferencias “no es consistente con la ausencia de diferencias entre muestras españolas y estadounidenses que presenta, por ejemplo, el BDI-II” (Sanz, 2007, p. 299).

Además, los manuales de las dos primeras versiones en castellano del MCMI-II proporcionan pocos datos sobre las propiedades psicométricas de las adaptaciones. Tan sólo ofrecen datos de consistencia interna en la 1ª y 2ª edición, mientras que la 3ª edición aporta información sobre la validez factorial y validez convergente-discriminante con el MMPI-2. Sin embargo, en ninguna de las ediciones en castellano del MCMI-II se proporcionan datos de validez diagnóstica.

Para la **adaptación española del MCMI-III**, se utilizaron procedimientos más parecidos a los de la versión original a la hora de elaborar los baremos de TB. En concreto, los baremos fueron obtenidos con una muestra de 964 pacientes (486 mujeres y 478 varones, con edades entre 18 y 76 años) de diferentes ciudades

españolas y contextos clínicos, evaluados por 105 profesionales familiarizados con el DSM-IV y la teoría de Millon. Las tasas de prevalencia de los trastorno fueron calculadas a partir de los diagnósticos realizados por dichos clínicos. El índice de consistencia interna de la escala de Depresión Mayor del MCMI-III en la muestra original fue de  $\alpha = 0,90$  y en la adaptación española fue de  $\alpha = 0,88$ , mientras que en el caso de la escala D, en la muestra española fue de  $\alpha = 0.87$ , mientras que en la original fue de  $\alpha = 0.88$ . Sin embargo, de nuevo no se obtuvieron o no se han proporcionado datos de validez diagnóstica del MCMI-III en su adaptación española.

Las Directrices de la Comisión Internacional de Tests son claras al respecto: “utilizar puntos de corte en la interpretación de las puntuaciones sólo cuando se disponga de datos empíricos sobre su validez” (Directriz 2.7.9.; ITC, 2000). En este sentido, cabe preguntarse si son generalizables a la versión-muestra española los datos de validez diagnóstica del MCMI-III original obtenidos en muestras de pacientes estadounidenses. Como se ha comentado, la generalización no se puede asumir automáticamente, sino que debe confirmarse empíricamente.

El MCMI está asociado a la recogida de información diagnóstica y sus propios autores señalan que ésta es uno de los objetivos principales del instrumento. De hecho, es uno de los usos para el que se está aplicando en la investigación y en la práctica clínica. Por ejemplo, en España uno de los principales manuales de evaluación psicológica recoge el MCMI como uno de los instrumentos diagnósticos de los trastornos mentales (Muñoz, 2003). Para que un instrumento de medida pueda ser utilizado con garantías dentro de la práctica clínica basada en la evidencia, se deben cumplir unos criterios de eficacia diagnóstica que en las escalas CC y D del MCMI-II y del MCMI-III no se conocen en sus adaptaciones españolas.



Sin embargo, debido a la ausencia de tales datos de validez diagnóstica, durante los últimos años, los usuarios de las adaptaciones españolas del MCMI-II o del MCMI-III no han podido utilizar sus escalas para el diagnóstico de los trastornos depresivos o lo han hecho a ciegas, sin saber con una mínima certeza sus propiedades como instrumento diagnóstico.



# Parte Empírica



# **Capítulo 5. OBJETIVOS DE LA INVESTIGACIÓN EMPÍRICA**



### 5.1. Justificación de la investigación empírica

Como se ha comentado en profundidad en el capítulo anterior, las adaptaciones españolas del MCMI-II y del MCMI-III carecen de estudios de validez diagnóstica. Uno de los objetivos principales de estos instrumentos es precisamente el diagnóstico de los trastornos mentales, entre ellos el trastorno depresivo mayor y el trastorno distímico, a partir de los puntos de corte propuestos en sus respectivos manuales (Millon, 1987, 1994; Millon et al., 1997). Sin embargo, en los manuales de las adaptaciones españolas de ambas versiones del inventario se recogen los datos de validez diagnóstica aportados en las ediciones originales con población estadounidense, pero no se recogen datos obtenidos con las versiones españolas en muestras de la población española (Millon, 1998, 1999, 2002, adaptación española de A. Ávila-Espada; Millon, Davis y Millon, 2007, adaptación española de V. Cardenal y M. P. Sánchez). Las Directrices de la Comisión Internacional de Tests son claras al respecto: se deben “utilizar puntos de corte en la interpretación de las puntuaciones sólo cuando se disponga de datos empíricos sobre su validez” (Directriz 2.7.9.; ITC, 2000). En este sentido, ¿son generalizables los datos de validez diagnóstica del MCMI-II y del MCMI-III originales en muestras de pacientes estadounidenses a la versión-población española? Como hemos comentado, la generalización no se puede asumir automáticamente, sino que debe confirmarse empíricamente.

Los autores del MCMI señalan que uno de los objetivos principales del cuestionario es la recogida de información diagnóstica y éste es uno de los usos más frecuentes para el que se está aplicando en la investigación y en la práctica clínica. Según Craig (1999, p. 398): “La evaluación de los trastornos afectivos ha sido el objetivo de muchos estudios realizados con el MCMI. El test tiene tres escalas de Distimia (D), Depresión Mayor (CC) y Bipolar (B) y claros puntos de

corte para el establecimiento de un diagnóstico de trastorno afectivo”. En España, uno de los principales manuales de evaluación psicológica recoge precisamente el MCMI como uno de los instrumentos diagnósticos de los trastornos mentales (Muñoz, 2003).

Para que un instrumento de medida pueda ser utilizado con garantías para fines diagnósticos dentro de la práctica clínica basada en la evidencia, se deben cumplir unos criterios de validez diagnóstica que en las escalas de CC y D del MCMI-II y del MCMI-III no están disponibles en sus adaptaciones españolas. Sin embargo, debido a la ausencia de tales datos de validez diagnóstica, durante los últimos años, los usuarios de las adaptaciones españolas del MCMI-II o del MCMI-III no han podido utilizar sus escalas para el diagnóstico de los trastornos depresivos o lo han hecho a ciegas, sin saber con la adecuada certeza sus propiedades como instrumento diagnóstico.

El diagnóstico de los trastornos mentales es actualmente uno de los retos más relevantes en el campo de la salud mental, con una importancia creciente. En la medida en la que, en los últimos 30 años, la actividad diagnóstica ha crecido espectacularmente y sistemas diagnósticos como el del DSM-IV están siendo ampliamente utilizados, debemos tener en cuenta que para realizar evaluaciones válidas, se debe ser riguroso en los métodos de asignación de etiquetas diagnósticas. En este sentido, es especialmente discutible el uso de los cuestionarios, escalas e inventarios como instrumentos diagnósticos en investigación y en la práctica clínica y, por tanto, precisamente por las dudas que suscita, la actividad diagnóstica basada en cuestionarios, escalas o inventarios debería hacerse con la máxima información posible sobre su validez.

Además, respecto a las versiones originales del MCMI existen dudas metodológicas sobre los estudios que han evaluado sus propiedades psicométricas



así como un déficit en los estadísticos publicados, de forma que la validez diagnóstica de este inventario ha sido objeto de una gran controversia (Hsu, 2002; Retzlaff, 1996; Grove y Vrieze, 2009).

## **5.2. Hipótesis de la investigación empírica**

En la presente tesis doctoral se tratará de comprobar la validez diagnóstica de las escalas de CC y D de las adaptaciones españolas del MCMI-II y del MCMI-III en relación con el diagnóstico de trastorno depresivo mayor y, de manera secundaria, en relación con el diagnóstico del conjunto de trastornos depresivos. Otros autores han encontrado también útil en el uso de la categoría “trastornos depresivos” para evaluar la validez diagnóstica del MCMI y, en general, de los instrumentos psicopatológicos, debido a que existe un solapamiento y redundancia entre los diagnósticos de trastorno depresivo mayor, distimia y trastorno depresivo no especificado (Guibeau y Choca, 2005). Además, utilizando la categoría “trastornos depresivos” se han encontrado resultados similares en el rendimiento diagnóstico de las escalas de CC y de D a los encontrados para cada trastorno por separado (Saulsman, 2011).

Por tanto, con este estudio, pretendemos responder a las siguientes preguntas:

1. ¿Cuál es la validez diagnóstica de las puntuaciones tasa base 75 y tasa base 85 de las escalas CC y D para el diagnóstico de trastorno depresivo mayor y de trastornos depresivos?
2. ¿Son mejores los índices de validez diagnóstica de otras puntuaciones directas o puntos de corte que los índices de las tasa base 75 y tasa base 85 de los baremos?

3. ¿Son mejores los índices de validez diagnóstica de las escalas del MCMI-II y MCMI-III que los de otros instrumentos para evaluar la depresión ampliamente utilizados como, por ejemplo, el BDI-II?

Las hipótesis de esta tesis doctoral son que las escalas CC y D de las adaptaciones españolas del MCMI-II y del MCMI-III mostrarán índices adecuados de validez diagnóstica en población española respecto al trastorno depresivo mayor y a los trastornos depresivos. Estas hipótesis estarían en la línea de lo encontrado en los estudios previos de rendimiento diagnóstico del MCMI-II (Millon, 2002) y del MCMI-III (Millon, 1997) en sus versiones originales en población estadounidense. Los resultados indicaban que dichas escalas mostraban índices de validez diagnóstica adecuados para poder ser utilizados tanto en investigación como en la práctica clínica en el diagnóstico de estos trastornos.

También se hipotetizó que un análisis de las curvas ROC podría revelar puntuaciones de corte más válidas de cara al diagnóstico que las propuestas en los manuales de las adaptaciones españolas (TB 75 y TB 85 de las muestras correspondientes de tipificación).

Finalmente, no se tenía una hipótesis concreta sobre la superioridad diagnóstica de las escalas CC y D del MCMI-II y del MCMI-III sobre el BDI-II, por lo que la presente investigación pretendía examinar también las características diagnósticas del BDI-II para los trastornos depresivos y compararlas de forma exploratoria con las del MCMI.

Con el objetivo de comprobar estas hipótesis se realizaron dos estudios independientes con el MCMI-II y el MCMI-III, respectivamente. En el caso del MCMI-II se trata de un estudio retrospectivo, de manera que no fue posible controlar algunas variables contaminadoras como, por ejemplo, el hecho de que los

evaluadores conocían las puntuaciones del MCMI-II antes de realizar el diagnóstico de los pacientes que sirvió como criterio de validez. Esta limitación fue corregida en el estudio sobre el MCMI-III, de forma que en el mismo los evaluadores debían asignar el diagnóstico a los pacientes sin conocer sus puntuaciones ni en el MCMI-III ni en el BDI-II. Además, en el primer estudio, los investigadores no habían utilizado necesariamente una medida estandarizada para la evaluación de la depresión, mientras que en el segundo estudio, los evaluadores asignaron las etiquetas diagnósticas después de la aplicación del módulo de trastornos del estado de ánimo de la SCID-I-VC. En la Tabla 5.1 se recogen las características más sobresalientes que distinguen los dos estudios de esta tesis doctoral.

Tabla 5.1. Características de los dos estudios de la presente investigación

	Test	Se utiliza una medida estandarizada para el diagnóstico (criterio de validez)	Se conocen los resultados del MCMI antes del diagnóstico
Estudio 1	MCMI-II	No necesariamente	Sí
Estudio 2	MCMI-III	Módulo de la SCID-I	No

**Capítulo 6. PRIMER ESTUDIO:  
VALIDEZ DIAGNÓSTICA DE  
LAS ESCALAS DE  
DEPRESIÓN MAYOR Y  
DISTIMIA DEL MCMI-II**



## 6.1. Introducción

Como se ha comentado con mayor detalle en el capítulo 4 de esta tesis doctoral, el MCMI, en sus diferentes versiones, es uno de los instrumentos más utilizados en psicología y, en especial, en psicología clínica. De hecho, un reciente estudio realizado con una muestra de 12 606 psicólogos profesionales procedentes de 17 países europeos revelaba que el MCMI ocupa la novena posición entre los instrumentos más utilizados por los psicólogos europeos (Evers et al., 2012). Teniendo en cuenta solamente los datos de ese estudio concernientes a España, los cuales fueron obtenidos a partir de una muestra de 3 126 psicólogos profesionales españoles, el MCMI ascendía a la tercera posición entre los instrumentos más utilizados por los psicólogos españoles, siendo el primero de los más utilizados por los psicólogos españoles que trabajan en el campo de la psicología clínica (Muñiz y Fernández-Hermida, 2010).

La adaptación española del MCMI-II (1998, adaptación española de A. Ávila-Espada) se publicó con posterioridad a la aparición en EE. UU. de la primera edición del MCMI-III (Millon, 1994). Puesto que hasta el año 2007 no se publicó en España el MCMI-III (Millon, Davis y Millon, 2007, adaptación española de V. Cardenal y M. P. Sánchez), es muy probable que muchos de los usuarios del MCMI en España continúen utilizando la segunda edición del inventario, el MCMI-II, en lugar de la más reciente, el MCMI-III. Efectivamente, en esta dirección parecen apuntar el hecho de que el número de estudios publicados en los dos últimos años (2011 y 2012) sobre el MCMI-II y el MCMI-III y recogidos en la base de datos bibliográfica Psycodoc es el mismo (6 sobre cada uno, aunque sólo 3 sobre el MCMI-III si no se tienen en cuenta los trabajos publicados por el grupo de investigación encargado de su adaptación).

A pesar de que se han publicado tres ediciones del MCMI-II en España (Millon, 1998, 1999, 2002, adaptación española de A. Ávila-Espada), no se ha realizado hasta la

fecha ningún estudio sobre la eficacia diagnóstica de sus escalas, siendo como es el diagnóstico clínico uno de sus principales objetivos (Millon, 1998). Por tanto, este estudio se centró en la utilización de las escalas del MCMI-II como instrumentos para la valoración diagnóstica, utilización que en estos momentos se está llevando a cabo en España sin que se conozcan las propiedades psicométricas de su versión española para este fin. En concreto, este estudio tenía como objetivo principal analizar la validez diagnóstica de sus escalas de Distimia (D) y Depresión Mayor (CC) en la detección del trastorno depresivo mayor y del conjunto de trastornos depresivos.

La relevancia de este objetivo no sólo radica en la posibilidad de poder contar con datos empíricos que sustenten la utilización diagnóstica de la versión española de dichas escalas del MCMI-II, sino también en la posibilidad de aportar datos sobre la validez diagnóstica de dichas escalas en general, ya que, tal y como se ha expuesto detalladamente en el capítulo 4 de esta tesis doctoral, es muy escasa la literatura científica al respecto. Efectivamente, además del estudio recogido en el manual de la propia versión original del MCMI-II (Millon, 2002), tan sólo se han podido localizar otros dos estudios publicados que hayan analizado la eficacia diagnóstica de las escalas D y CC del MCMI-II para identificar los trastornos depresivos (Piersma, 1991; Wetzler y Marlowe, 1993). Por otro lado, los resultados de estos estudios, que se resumen en la Tabla 6.1, no han sido totalmente consistentes, de manera que, por ejemplo, los índices de eficacia diagnóstica hallados por Wetzler y Marlowe (1993) no solo son notablemente más pobres que los informados por Millon (2002), sino que arrojan serias dudas sobre la validez diagnóstica de las escalas CC y D del MCMI-II al obtener sensibilidades e índices de eficacia global muy por debajo del valor de 70%, que es considerado el mínimo aceptable en un instrumento diagnóstico (Lenderink et al., 2012; Sheehan et al., 1998).

**Tabla 6.1.** Índices de validez diagnóstica de las escalas de Depresión Mayor y Distimia del MCMI-II para las puntuaciones de tasa base 75 y 85 en diversos estudios con muestras heterogéneas de pacientes con trastornos psicológicos

Índices	Millon (2002) <sup>a</sup>				Piersma (1991) <sup>b</sup>				Wetzler y Marlowe (1993) <sup>c</sup>			
	Depresión		Distimia		Depresión		Distimia		Depresión		Distimia	
	TB7 5	TB 85	TB 75	TB 85	TB 75	TB 85	TB 75	TB 85	TB 75	TB 85	TB 75	TB 85
Prevalencia	14	6	46	29	77	—	77	—	55	55	55	55
Sensibilidad	68	47	81	76	61	—	86	—	18	10	57	33
Especificidad	96	97	83	88	52	—	32	—	86	92	70	86
VPP	73	50	80	72	81	—	81	—	47	46	57	52
VPN	95	97	84	90	28	—	40	—	61	60	69	74
Eficacia	92	94	82	84	59	—	73	—	49	47	63	57

*Nota.* VPP: valor predictivo positivo. VPN: valor predictivo negativo. Eficacia: índice de eficacia global o porcentaje de pacientes correctamente clasificados (también denominado precisión, índice de validez o de eficiencia, o poder o valor diagnóstico total).

<sup>a</sup>Eficacia para el diagnóstico del trastorno depresivo mayor en el caso de la escala de Depresión Mayor y para el diagnóstico del trastorno distímico en el caso de la escala de Distimia.

<sup>b</sup>Eficacia para el diagnóstico del trastorno depresivo mayor. La muestra de este estudio no era heterogénea, sino que todos los pacientes tenían trastornos depresivos (trastorno depresivo mayor, distímico, adaptativo con estado de ánimo deprimido y depresivo no especificado).

<sup>c</sup>Eficacia para el diagnóstico de cualquier trastorno depresivo.

En resumen, para aportar más datos a la escasa literatura científica sobre la validez diagnóstica de las escalas de Depresión Mayor (CC) y Distimia (D) del MCMI-II y realizar el primer estudio sobre dicha validez en el caso de la versión española del MCMI-II, se llevó a cabo un estudio retrospectivo sobre los datos proporcionados por una



muestra de pacientes españoles con diversos trastornos psicológicos a los que se les había administrado el MCMI-II y de los que se disponía del diagnóstico DSM-IV asignado por los psicólogos que habían actuado como sus terapeutas tras mantener con ellos varias sesiones de evaluación. Este diagnóstico sirvió como criterio de referencia para evaluar la eficacia diagnóstica de las escalas CC y D del MCMI-II respecto a la capacidad para detectar, en primer lugar, pacientes con trastorno depresivo mayor y, en segundo lugar, pacientes con todo tipo de trastornos depresivos, incluyendo en este último grupo no sólo los dos tipos principales considerados por el DSM-IV (trastorno depresivo mayor y trastorno distímico), sino también todas las categorías que el DSM-IV considera residuales y relacionadas con la depresión (trastorno adaptativo con estado de ánimo depresivo, trastorno adaptativo mixto, duelo y trastorno depresivo no especificado).

Por otro lado, Ganellen (1996), tras revisar los estudios sobre la validez diagnóstica del MCMI-II y constatar que el método de asignación de los participantes al grupo de comparación no deprimido utilizado en dichos estudios podía disminuir los valores de especificidad e incrementar la tasa de falsos positivos, recomendaba “que los investigadores que estudian una muestra de pacientes con una forma relativamente pura de un trastorno en particular, tales como pacientes con depresión, excluyan del grupo de comparación a los pacientes que tienen otras formas del mismo trastorno, tales como pacientes con distimia o un trastorno adaptativo con estado de ánimo deprimido” (Ganellen, 1996, p. 229). Siguiendo, pues, la recomendación de Ganellen (1996), en el presente estudio también se analizó la eficacia diagnóstica de las escalas CC y D del MCMI-II para identificar a los pacientes con trastorno depresivo mayor frente a los pacientes sin depresión, excluyendo, por tanto, de este último grupo a los pacientes con trastorno distímico, trastorno adaptativo con estado de ánimo depresivo, duelo y trastorno depresivo no especificado, así como a los pacientes con un diagnóstico secundario de trastorno depresivo.

Finalmente, como un objetivo secundario, el presente estudio también pretendía comparar la eficacia diagnóstica del MCMI-II frente al Inventario de Depresión de Beck-II (BDI-II). Aunque el propósito original del BDI-II no es realizar diagnósticos clínicos, en muchas ocasiones ha sido utilizado como instrumento de cribado para detectar pacientes con un posible trastorno depresivo que, posteriormente, debería ser confirmado mediante una entrevista clínica o cualquier otro instrumento diagnóstico más apropiado (p. ej., una entrevista semiestructurada o estructurada). Por otro lado, el BDI, en sus diferentes versiones, incluyendo el BDI-II, es un instrumento ampliamente utilizado para evaluar la depresión, por lo que parecía un candidato muy oportuno para compararlo con las escalas del MCMI-II. De hecho, en los estudios europeo y español que se mencionaban anteriormente (Evers et al., 2012; Muñoz y Fernández-Hermida, 2010), el BDI se situaba en el sexto lugar de los tests más utilizados por los psicólogos profesionales europeos, en el quinto lugar de los más utilizados por los psicólogos profesionales españoles y en el cuarto lugar de los más utilizados por los psicólogos españoles que trabajan en el área de la psicología clínica.

## **6.2. Método**

### **6.2.1. Participantes**

En este estudio sobre las escalas de los trastornos depresivos del MCMI-II participaron inicialmente 194 pacientes con trastornos psicopatológicos atendidos en la Clínica Universitaria de Psicología de la Facultad de Psicología de la Universidad Complutense de Madrid (UCM). La muestra estaba compuesta por hombres y mujeres mayores de 18 años, a los que se les había aplicado el MCMI-II y cuyas puntuaciones en los índices de validez del MCMI-II no cumplían las condiciones de invalidación de la prueba establecidas en el manual del MCMI-II (Millon, 2002, adaptación española de A. Ávila-Espada), de manera que 4 pacientes no se incluyeron en la muestra por no cumplir

los criterios de exclusión del MCMI-II. Además, 9 pacientes (4.7% de la muestra inicial), fueron eliminados del análisis sobre la eficacia diagnóstica de las escalas de los trastornos depresivos del MCMI-II debido a que los diagnósticos no señalaban claramente si en ese momento tenían un episodio depresivo (3 pacientes diagnosticados con un trastorno bipolar en remisión, sin la especificación sobre si el último episodio era depresivo o maniaco; 2 pacientes con un trastorno adaptativo mixto emoción-conducta, sin la especificación sobre si la respuesta emocional era de tipo ansiosa o depresiva, y 4 pacientes con diagnóstico aplazado). Por tanto, en la muestra definitiva se incluyeron 181 pacientes. La edad media de la muestra fue de 32.23 años, con una desviación típica de 11.59. En las Tablas 6.2, 6.3, 6.4, 6.5, 6.6 y 6.7 se describen las principales características sociodemográficas (sexo, estado civil, nivel de estudios y situación laboral) y clínicas (demanda asistencial y diagnósticos psicopatológicos según el DSM-IV) de la muestra.

**Tabla 6.2.** Sexo en la muestra del estudio 1

Sexo	Frecuencia	Porcentaje
Varones	60	33.1%
Mujeres	121	66.9%
Total	181	100%

**Tabla 6.3.** Estado civil en la muestra del estudio 1

Estado Civil	Frecuencia	Porcentaje
Soltero	107	59.1
Casado/convivencia estable con pareja	41	22.7
Viudo	16	8.8
Separado/divorciado	3	1.7
Total	168	92.8
Sin datos	13	7.2

**Tabla 6.4.** Nivel de estudios en la muestra del estudio 1

<b>Nivel de Estudios</b>	<b>Frecuencia</b>	<b>Porcentaje</b>
Estudios primarios	5	2.8
EGB	9	5
Bachillerato	67	37
Universitarios	24	13.3
Otros	63	34.8
Total	168	92.8
Sin datos	13	7.2

**Tabla 6.5.** Profesión/situación laboral en la muestra del estudio 1

<b>Profesión/Situación laboral</b>	<b>Frecuencia</b>	<b>Porcentaje</b>
Obrero sin cualificar	3	1.7
Obrero cualificado	4	2.2
Sector servicios	11	6.1
Personal servicios administrativos	17	9.4
Empresario	2	1.1
Directivo/gerente	4	2.2
Profesional/técnico	34	18.8
Ama de casa	2	1.1
Estudiante	78	43.1
Jubilado debido a la edad	2	1.1
Jubilado por enfermedad o minusvalía	1	0.6
Profesional de las fuerzas armadas	1	0.6
Parado	5	2.8
Otros	2	1.1
Total	166	91.7
Sin datos	15	8.3

**Tabla 6.6.** Demanda asistencial en la muestra del estudio 1

<b>Demanda asistencial</b>	<b>Frecuencia</b>	<b>Porcentaje</b>
Problemas de ansiedad	37	20.4
Problemas del estado de ánimo	28	15.5
Dificultades en las relaciones sociales (habilidades sociales)	5	2.8
Problemas de pareja	10	5.5
Problemas sexuales	3	1.7
Problemas de autoestima	5	2.8
Problemas de estrés	2	1.1
Problemas de sueño	1	0.6
Problemas infantiles	2	1.1
Problemas de adicciones	1	0.6
Problemas de alimentación	5	2.8
Síntomas psicóticos	3	1.7
Otros	28	15.5
Varios de los anteriores	48	26.5
Demanda inespecífica	2	1.1
Total	180	99.4
Sin datos	1	0.6

**Tabla 6.7.** Diagnósticos según el DSM-IV en la muestra del estudio 1

<b>Diagnóstico</b>	<b>Frecuencia</b>	<b>Porcentaje</b>
Trastorno depresivo mayor	24	13.3
Trastorno distímico	10	5.5
Trastorno depresivo no especificado	3	1.7
Trastorno de angustia sin agorafobia	4	2.2
Trastorno de angustia con agorafobia	6	3.3
Agorafobia sin trastorno de angustia	1	0.6
Fobia específica	4	2.2
Fobia social	14	7.7
Trastorno obsesivo compulsivo	5	2.8
Trastorno por estrés postraumático	3	1.7

Diagnóstico	Frecuencia	Porcentaje
Trastorno de ansiedad generalizado	7	3.9
Trastorno de ansiedad no especificado	18	9.9
Trastorno adaptativo con estado de ánimo depresivo	9	5
Trastorno adaptativo con ansiedad	2	1.1
Trastorno adaptativo mixto ansiedad-ánimo depresivo	10	5.5
Anorexia	4	2.2
Bulimia	6	3.3
Trastorno de la conducta alimentaria no especificado	4	2.2
Trastornos relacionados con consumo de sustancias	1	0.6
Duelo	1	0.6
Esquizofrenia y otros trastornos psicóticos	4	2.2
Problemas conyugales	13	7.2
Trastornos del sueño	1	0.6
Trastornos somatomorfos	2	1.1
Trastornos sexuales	2	1.1
Otros diagnósticos	21	11.5
Sin diagnóstico en el Eje I	2	1.1
Total	181	100

La Clínica Universitaria de Psicología de la UCM es un centro que presta atención psicológica ambulatoria a estudiantes y personal de la UCM, pero también a personas ajenas a la universidad (García-Vera y Sanz, 2009), y, quizás por esa razón, el perfil de pacientes y problemas clínicos que atiende es comparable en muchos aspectos al que se puede encontrar en centros privados españoles de psicología clínica (Estupiñá, Labrador y García-Vera, 2012). No obstante, en la muestra de pacientes del presente estudio se obtuvo una puntuación media en el Inventario de Depresión de Beck-II (BDI-II) de 19.83 ( $DT = 10.65$ ), ligeramente inferior a la que se ha encontrado en otras investigaciones con muestras de pacientes psicopatológicos. Así, en una revisión reciente de las características psicométricas del BDI-II en muestras de pacientes psicopatológicos de

diversos países, se calculó una puntuación media de 22.8 ( $DT = 11.9$ ) promediando de forma ponderada los resultados de 10 estudios que totalizaban 2 978 pacientes (Sanz y García-Vera, 2009). En este mismo sentido, las medianas de las puntuaciones directas en la escala de Depresión Mayor (CC) del MCMI-II de la presente muestra son inferiores a las TB 60 de la muestra de tipificación de la adaptación española del MCMI-II, las cuales, por definición, representan la mediana de esa muestra psicopatológica de tipificación. Así, las medianas de la escala de Depresión Mayor (CC) del MCMI-II en los pacientes varones y mujeres de esta investigación fueron 16 y 23, respectivamente, mientras que las TB 60 de la muestra de tipificación española de pacientes varones y mujeres fueron 18 y 24-25, respectivamente (Millon, 2002, adaptación española de A. Ávila-Espada).

Como se puede observar en la Tabla 6.7, en esta muestra la prevalencia del trastorno depresivo mayor fue del 13.3% y la de la distimia del 5.5%. En línea con los datos anteriores, estas cifras fueron menores a las encontradas en otros estudios sobre la prevalencia de los trastornos mentales en centros de salud mental en España, que informan de prevalencias de los trastornos depresivos superiores al 20% (de Francisco y Goenaga, 2002; Valero y Ruiz, 2003). Por otro lado, en la muestra de este estudio, se encontró una prevalencia del trastorno depresivo no especificado del 1.7%, del trastorno adaptativo con estado de ánimo depresivo del 5%, del trastorno adaptativo mixto del 5.5% y del duelo del 0.6%, por lo que la prevalencia del conjunto más general de trastornos depresivos fue del 31.6%.

### **6.2.2. Instrumentos**

En la presente investigación se utilizaron los siguientes instrumentos:

1. **MCMI-II** (Millon, 1987) **en la tercera edición de su adaptación española** (Millon, 2002, adaptación española de A. Ávila-Espada). El MCMI-II es un instrumento de autoinforme que consta de 175 ítems de respuesta verdadero-falso y que permite obtener medidas en 2 escalas de control, 10 escalas básicas de patrones clínicos de personalidad, 3 escalas de patología graves de personalidad, 6 síndromes clínicos de gravedad moderada y 3 síndromes de gravedad más extrema (véase la descripción de las escalas en el Anexo 5). De todas las escalas del MCMI-II, en el presente estudio se utilizaron las escalas de Distimia (D) y de Depresión Mayor (CC) y dos escalas que señalan hasta qué punto el perfil del MCMI-II se considera válido o inválido. Estas dos escalas son las siguientes (Millon, 2002):

- **Validez** (escala V): El índice de validez está compuesto por cuatro frases muy poco prevalentes en los sujetos. Los sujetos que contestan a estas afirmaciones de una manera poco esperable puntúan alto en esta escala y las correcciones aparecen como no válidas o deben interpretarse con cautela.

- **Sinceridad** (escala X): Indica si el paciente se inclina a ser reservado y reticente, o por el contrario es franco y revelador. Se obtiene de la suma de las puntuaciones en las escalas básicas de personalidad.

Siguiendo los criterios establecidos en el manual del MCMI-II (Millon, 2002, adaptación española de A. Ávila-Espada), en el presente estudio se incluyeron aquellos pacientes que puntuaron  $< 2$  en la escala V y  $> 144$  y  $< 590$  en la escala X. Los pacientes con puntuaciones que no cumplían estos criterios quedaron excluidos. Dado que los baremos de la primera y segunda edición de la versión española del MCMI-II difieren de aquellos de la tercera edición (véase Sanz, 1997), es importante subrayar que para el cálculo de las puntuaciones TB se han utilizado los baremos de corrección de la tercera edición en castellano del MCMI-II (Millon, 2002, adaptación



española de A. Ávila-Espada). La adaptación española de las escalas D y CC del MCMI-II presenta índices adecuados de fiabilidad de consistencia interna y de validez convergente (Millon, 2002, adaptación española de A. Ávila-Espada).

2. **Inventario de Depresión de Beck, segunda edición (BDI-II;** Beck et al., 1996) **en su adaptación española** (Sanz et al., 2005; Sanz y García-Vera, 2013; Sanz, Navarro y Vázquez, 2003; Sanz, Perdigón y Vázquez, 2003) que recientemente ha sido publicada por *Pearson Talent and Assessment España* (Beck, Steer y Brown, 2011a, adaptación de J. Sanz y C. Vázquez). Este cuestionario de autoinforme está compuesto por 21 ítems diseñados para evaluar la gravedad de la sintomatología depresiva en adultos y adolescentes con una edad mínima de 13 años. En cada uno de los ítems la persona tiene que elegir, entre un conjunto de cuatro alternativas ordenadas de menor (0) a mayor gravedad (3) (salvo dos ítems que tienen siete alternativas, pero que se puntúan también entre 0 y 3), la frase que mejor describe su estado durante las últimas dos semanas incluyendo el día en que completa el inventario. El BDI-II es uno de los instrumentos más utilizados para evaluar la sintomatología depresiva, así como el cambio en dicha sintomatología durante el proceso terapéutico y en el postratamiento. Además, la adaptación española muestra buenos datos tanto de fiabilidad de consistencia interna como de validez convergente, discriminante, diagnóstica y factorial, y tanto en muestras españolas de adultos de la población general (Sanz, Perdigón y Vázquez, 2003) y de estudiantes universitarios (Sanz et al., 2003) como de pacientes con trastornos psicológicos (Sanz et al., 2005; Sanz y García-Vera, 2013).
3. **Listado para el Episodio Depresivo (LED;** Sanz et al., 2005). Este es un instrumento de heteroevaluación para medir la presencia de síntomas depresivos y valorar la gravedad de la depresión en pacientes con trastornos psicológicos. Consta

de dos preguntas abiertas y de 12 preguntas con tres opciones de respuesta [“Sí”, “No” y “¿?” (sin información suficiente-no estoy seguro)] que cubren los criterios diagnósticos del episodio depresivo mayor del DSM-IV (APA, 1994). El LED fue diseñado para que un clínico, después de una entrevista de al menos 45 minutos con el paciente, coteje dichos criterios y pueda determinar la presencia o ausencia de un diagnóstico de trastorno depresivo. Además, el instrumento incluye una última pregunta en la que el clínico debe estimar la gravedad de los síntomas depresivos del paciente (aunque no se correspondan con ningún diagnóstico) en una escala de tipo Likert de cuatro puntos (“sin depresión/depresión mínima”, “depresión leve”, “depresión moderada” y “depresión grave”). En muestras de pacientes con trastornos psicológicos, el LED muestra índices adecuados de fiabilidad de consistencia interna (entre 0.63 y 0.69) e índices entre adecuados y excelentes de validez convergente con el BDI-II (entre 0.44 y 0.66) (Sanz, 2013).

### **6.2.3. Procedimiento**

En la Clínica Universitaria de Psicología de la UCM se llevó a cabo desde el año 2006 a mayo de 2008 un protocolo de evaluación que incluía entre otros instrumentos la aplicación del MCMI-II, el BDI-II y el LED. Durante la primera o segunda sesión (de aproximadamente 1 hora de duración cada una) los evaluadores aplicaban los cuestionarios. Al finalizar esta consulta, los evaluadores revisaban las preguntas de chequeo del LED y rellenaban a continuación una ficha con el diagnóstico del sujeto siguiendo el sistema de clasificación de trastornos mentales del DSM-IV (véase el Anexo 6). Los diagnósticos utilizados en este estudio se obtuvieron de esa ficha, pero, con el objetivo de tener la muestra más amplia posible, en el caso de que esa ficha no estuviera cumplimentada, se utilizó el diagnóstico que figuraba en la historia del paciente realizado

tras el proceso de evaluación completo, de modo que el 24% de los diagnósticos de la muestra se obtuvieron de la historia del paciente.

En la presente investigación se revisaron todos los historiales de los pacientes mayores de 18 años que acudieron a la Clínica Universitaria de Psicología durante dicho periodo de tiempo, con un diagnóstico definitivo asignado por parte de los evaluadores, que tuvieran cumplimentados tanto el MCMI-II como el BDI-II y un diagnóstico en la ficha diagnóstica o en el informe realizado por el clínico tras el periodo de evaluación.

En el proceso de evaluación, los psicólogos tenían libertad para elegir cualquier otro procedimiento de recogida de información como entrevistas de contraste con familiares, entrevistas estructuradas, libres o semiestructuradas, así como cuestionarios, escalas o inventarios adicionales. Además los psicólogos seleccionaban el orden de aplicación de los instrumentos.

Los evaluadores conocían generalmente las puntuaciones de los pacientes tanto del MCMI-II como del BDI-II antes de realizar el diagnóstico del paciente, de manera que no se pudo descartar la posibilidad de sesgos o de contaminación cruzada entre las puntuaciones de estos inventarios y los diagnósticos realizados por los psicólogos.

#### **6.2.4. Evaluadores**

Cada paciente fue evaluado únicamente por el psicólogo al que fue asignado como terapeuta en la Clínica. Todos los terapeutas fueron psicólogos becarios-residentes de la Clínica Universitaria de Psicología de la UCM, con formación de postgrado en psicología clínica y de la salud y como mínimo con 2 años de experiencia clínica. Todos ellos, además, fueron entrenados en la aplicación de los instrumentos de evaluación descritos y en los pasos a seguir para la correcta aplicación de los mismos.

### 6.2.5. Diseño y análisis de datos

El diseño de la investigación es un diseño retrospectivo instrumental, de análisis de las características diagnósticas de un instrumento.

Existen varios estadísticos para el análisis de la validez diagnóstica de los instrumentos clínicos, por lo que hay discrepancias entre los diferentes investigadores en cuanto a los estadísticos a publicar y analizar (Hsu, 2002). También existen discrepancias con respecto a la metodología que hay que utilizar para alcanzar los puntos de corte diagnóstico de un cuestionario y distintas posturas acerca de la importancia diferencial que se debe otorgar a los falsos positivos y a los falsos negativos en la tarea diagnóstica; pero lo fundamental es que los clínicos que deben interpretar los resultados diagnósticos tengan en cuenta la naturaleza de los índices estadísticos utilizados para probar la validez diagnóstica de un instrumento. Es decir, que sea cual sea el procedimiento que se siga para determinar la validez diagnóstica, se debe informar explícitamente sobre qué sistema se ha elegido y proporcionar información suficiente a los usuarios e investigadores para valorar su adecuación para sus propósitos. De lo contrario, existirá una fuente de contaminación que puede invalidar desde su base tanto las evaluaciones de los clínicos como las futuras investigaciones (Hsu, 2002).

Los estadísticos más utilizados para estudiar la validez diagnóstica de los instrumentos, y que se utilizaron en la presente investigación, son los siguientes:

**1) Sensibilidad:** señala la probabilidad de que un paciente diagnosticado con un trastorno obtenga un resultado positivo en el test para ese trastorno. Es la proporción (o el porcentaje) de pacientes que tienen efectivamente el diagnóstico clave que puntúan por encima o igual al punto de corte. Se calcula dividiendo el número de pacientes diagnosticados correctamente por el instrumento entre el número total de pacientes que

efectivamente tienen dicho trastorno en la muestra. En general, una prueba que se utilice a modo de cribado o *screening* deberá tener una gran sensibilidad, para lograr detectar los posibles casos dentro de la muestra. El índice de sensibilidad puede tomar valores entre 0 o 1, o entre 0% y 100% cuando se expresa en porcentajes, y, aunque, tal y como se ha dicho antes, los objetivos para los que va a ser usado el instrumento determinan qué valores de sensibilidad se consideran adecuados, es habitual considerar aceptables los iguales o mayores de 0.70 o 70% (Lenderink et al., 2012; Sheehan et al., 1998). Por ejemplo, en el trabajo de Lenderink et al. (2012), tanto para la sensibilidad como para la especificidad, un “valor > 85% fue considerado alto, 70–85% fue considerado moderado, y < 70% fue considerado bajo” (p. 232).

**2) Especificidad:** es el porcentaje de pacientes que puntúan por debajo del punto de corte en el test sin la presencia del diagnóstico clave; es decir, la probabilidad de detectar correctamente a un individuo que no tiene el trastorno. Se calcula dividiendo el número de pacientes sin diagnóstico que ha sido detectado correctamente por el cuestionario, entre el número total de pacientes que efectivamente no tienen dicho trastorno en la muestra. Es un estadístico apropiado sobre todo para pruebas confirmatorias de un diagnóstico o en psicología forense.

La especificidad, al igual que la sensibilidad, es un estadístico sencillo de calcular e intuitivo en su interpretación, que puede tomar valores entre 0 o 1, o entre 0% y 100% cuando se expresa en porcentajes, y que es considerado aceptable cuando alcanza valores iguales o mayores de 0.70 o 70% (Lenderink et al., 2012), aunque, por supuesto, los objetivos para los que va a ser usado el instrumento determinan qué valores de especificidad serían los más adecuados.

**3) Poder o valor predictivo positivo:** es el porcentaje de pacientes que puntúan por encima del punto de corte en el test y que efectivamente tienen el trastorno. Por

ejemplo, en el caso concreto de que un paciente puntúe por encima de la TB 85 en la escala CC del MCMI-III, el poder predictivo positivo de la TB 85 indicaría la probabilidad de que esa persona tuviera efectivamente el diagnóstico de depresión mayor. No se ha encontrado ninguna propuesta concreta sobre cuál debería ser el poder o valor predictivo positivo mínimo de un instrumento para que éste fuera considerado aceptable como instrumento diagnóstico, pero cabe suponer que al menos dicho índice debería mayor de 50%, ya que entonces indicaría que la probabilidad que tiene un paciente que haya superado el punto de corte en un test de tener efectivamente el trastorno en cuestión es mayor que la probabilidad que supondría lanzar simplemente una moneda para decidir si un paciente padece o no dicho trastorno.

**4) Poder o valor predictivo negativo:** porcentaje de pacientes con puntuación menor que el punto de corte sin el diagnóstico clave. Es decir, dado que una persona puntúa por debajo del punto de corte, cuál es la probabilidad de que efectivamente no cumpla los criterios para ese diagnóstico. Al igual que sucede con el poder predictivo positivo, no se ha encontrado ninguna propuesta concreta sobre cuál debería ser el poder predictivo negativo mínimo de un instrumento para que éste fuera considerado aceptable como instrumento diagnóstico, pero también con este índice cabría suponer que al menos debería ser mayor de 50%, ya que entonces indicaría que la probabilidad que tiene un paciente que puntúe por debajo del punto de corte en un test de no padecer efectivamente el trastorno en cuestión es mayor que la probabilidad que supondría lanzar simplemente una moneda para decidir si un paciente padece o no dicho trastorno.

**5) Índice kappa:** Es el grado de acuerdo entre el instrumento y el criterio de referencia, en el caso de este estudio, el diagnóstico proporcionado por el clínico tras el proceso de evaluación. Es un estadístico que permite valorar la variabilidad entre dos fuentes de medida independientes, siempre que la variable sea categorial. Se utiliza para

comparar la proporción de sujetos clasificados correctamente por el instrumento con los que se clasificarían al azar (Hsu, 2002). Índices negativos implican un grado de acuerdo menor del que se obtendría al azar. A la hora de interpretar el valor de kappa es útil disponer de una escala de referencia. Altman (1991) señala que, a pesar de su arbitrariedad, la más ampliamente utilizada es la que se recoge en la Tabla 6.8.

**Tabla 6.8.** Propuesta de interpretación del índice kappa (Altman, 1991)

<b>Valor de kappa</b>	<b>Fuerza de la concordancia</b>
< 0.20	Pobre
0.21 – 0.40	Débil
0.41 – 0.60	Moderada
0.61 – 0.80	Buena
0.81 – 1.00	Muy buena

El índice kappa, además de considerarse un índice de validez diagnóstica cuando se compara una medida o instrumento respecto a un criterio de referencia, también se considera el principal estadístico para el análisis de la fiabilidad diagnóstica interjueces o interevaluadores. Así, Kraemer, Kupfer, Clarke, Narrow y Regier (2012) han destacado recientemente que el índice kappa es el estadístico más utilizado para el análisis de la fiabilidad de los diagnósticos clínicos, y, para este propósito, se han recomendado kappas superiores a 0.70 (Joiner, Walker, Pettit, Pérez y Cukrowicz, 2005) para considerar un instrumento como fiable.

Los estadísticos de poder predictivo e índice kappa proporcionan información fundamental para la toma de decisiones de los clínicos, pero tienen una dificultad: se ven

afectados por la prevalencia del trastorno, por lo que otorgan valores menores a los trastornos menos prevalentes (Hsu, 2002). Además, se calculan a partir de un punto de corte determinado por el investigador y por tanto son sensibles al cambio en el punto de corte seleccionado.

**6) Índice de Youden.** Este estadístico refleja la relación entre la sensibilidad y la especificidad. Se calcula sumando la sensibilidad y la especificidad y restándole 1 al resultado de la suma. Cuanto más se acerque el índice de Youden a 1, mejor se considera el instrumento diagnóstico. Se puede calcular para los distintos puntos de corte del instrumento, de manera que ayuda a seleccionar el punto que tiene una mejor combinación entre sensibilidad y especificidad. Puesto que los objetivos para los que va a ser usado el instrumento determinan qué valores de sensibilidad y especificidad serían los más adecuados, es obvio que dichos objetivos también determinan qué valores del índice de Youden serían los más adecuados. De hecho, no se han encontrado en la literatura científica propuestas de valores de referencia para este índice. No obstante, puesto que habitualmente se suele considerar un valor igual o mayor de 0.70 (o 70%) como índices aceptables de sensibilidad y especificidad (Lenderink et al., 2012; Sheehan et al., 1998), tendría sentido considerar como aceptables índices de Youden iguales o mayores a los obtenidos con esos estándares de sensibilidad y especificidad, es decir, índices de Youden iguales o mayores que 0.40 (o que 40, cuando la sensibilidad y la especificidad se expresan en porcentajes).

**7) Eficacia global** (también denominado índice de validez, índice de eficiencia o valor o poder diagnóstico total). Es un estadístico utilizado con frecuencia para describir la eficacia diagnóstica de las pruebas y se refiere al porcentaje de pacientes correctamente clasificados.



**8) Área bajo la curva ROC.** Es un estadístico que refleja la relación entre la sensibilidad del instrumento y su especificidad para cada una de sus puntuaciones. Es una herramienta desarrollada a partir de la Segunda Guerra Mundial para mejorar la detección de las señales de radar y se utiliza en un amplio abanico de contextos en la actualidad (Fan, Upadhye y Worster, 2006). En relación con la eficacia diagnóstica de los instrumentos diagnósticos, el área bajo la curva de ROC refleja la probabilidad de que una persona elegida al azar con el trastorno, tenga una puntuación que exceda la de otra persona seleccionada sin ese diagnóstico. Toma valores desde 0 a 1 (Fawcett, 2004). En el caso de que el área de la curva de ROC sea 0.5, este valor señala que la capacidad del instrumento para identificar correctamente un caso (ya sea como positivo o negativo para la característica analizada) es igual a la de una clasificación realizada en base al azar, mientras que un área con un valor de 1 indica que el instrumento logra una clasificación perfecta del conjunto de casos analizados (Swets, 1988).

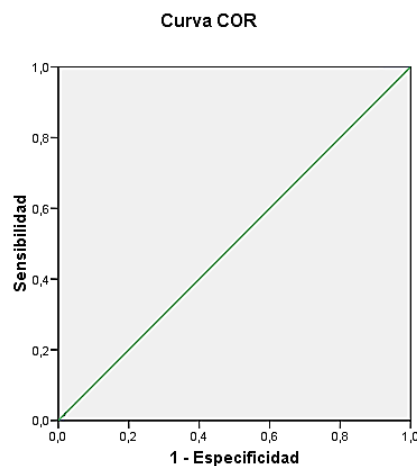
Una importante ventaja de este estadístico es que otorga dichos valores con independencia de la prevalencia del trastorno. Esta ventaja es importante puesto que la prevalencia de un trastorno puede variar en distintas poblaciones y en distintos momentos temporales. Pero este estadístico sí es sensible con respecto al uso para el que esté siendo utilizado el instrumento, así como a las circunstancias y población en la que se aplique.

El gráfico del área bajo la curva ROC es una representación de la sensibilidad y de la tasa de falsos negativos ( $1 - \text{especificidad}$ ), que nos permite visualizar de manera rápida e intuitiva hasta qué punto el instrumento nos sirve para diferenciar entre la población con el trastorno de la que no lo tiene. Como se puede ver en la Figura 6.1, en la curva ROC la sensibilidad se distribuye en el Eje Y y la tasa de falsos negativos ( $1 - \text{especificidad}$ ) en el Eje X. En este sentido, el instrumento perfecto sería el que tuviera un

punto de corte con una sensibilidad igual a 1 y una especificidad igual a 1 (porque se detectarían todos los casos que realmente tuvieran el trastorno y además ningún paciente sin trastorno sería detectado como si lo tuviera). Una especificidad de 1 supondría que la tasa de falsos negativos ( $1 - \text{especificidad}$ ) sería 0, por lo que en la representación gráfica de la Figura 6.1, las coordenadas que indicarían un instrumento perfecto se situarían arriba a la izquierda (1, 0). Cuando más cerca de este punto esté la curva de ROC mayor será su área, lo que indicaría que el instrumento muestra una mejor eficacia diagnóstica. “De manera informal, un punto en el espacio ROC es mejor que otro cuanto más situado al noroeste se encuentre con respecto al primero” (Fawcett, 2004, p. 4).

Ese mismo autor señala que cuanto más cerca esté la curva a la izquierda (eje X), el test sería más conservador en el sentido de que los valores que detecta como positivos tienen más evidencia (especificidad), aunque no detecte todos los verdaderamente positivos (sensibilidad). Por el contrario, los puntos situados arriba a la derecha están relacionados con puntuaciones que tienen mayor probabilidad de detectar los casos positivos, pero pueden detectar como positivos casos que en realidad no lo son.

**Figura 6.1.** Plano sobre el que se sitúa el gráfico del área bajo la curva ROC (denominada curva COR en el paquete estadístico SPSS)



Los segmentos diagonales son producidos por los empates.

Como se ha comentado, cuando el área de la curva ROC es 0.5 se puede considerar que el instrumento de medida es igual de hábil que el azar para detectar sujetos con trastornos (en la Figura 6.1, el azar se correspondería con la línea recta que va desde las coordenadas 0, 0 a 1, 1, es decir, con la diagonal dibujada en verde). Si una puntuación de un test, por ejemplo, se sitúa en las coordenadas (0.9, 0.9), en realidad está señalando que el instrumento puede detectar una mayoría de personas con la patología, pero no detecta correctamente los pacientes sin patología. Cualquier test que tenga una gráfica por debajo de la línea diagonal de la Figura 6.1 (en verde), mostraría una eficacia diagnóstica más baja que el azar (Fan et al., 2006).

Swets (1988, p. 1292) considera que valores del área bajo la curva ROC “entre 0.0 y 0.70 aproximadamente representan una exactitud bastante baja—la proporción de verdaderos positivos no es mucho mayor que la proporción de falsos positivos en cualquier punto a lo largo de la curva. Valores del [área bajo la curva] entre aproximadamente 0.70 y 0.90 representan exactitudes que son útiles para algunos propósitos, y valores más altos representan una exactitud bastante alta”. Los criterios de Swets (1988) para interpretar los valores del área bajo la curva han sido adoptados por otros investigadores. Por ejemplo, tanto Fischer, Bachmann y Jaeschke (2003) como Streiner y Cairney (2007) consideran que un test con un área bajo la curva entre 0.50 y 0.70 tiene exactitud baja, entre 0.70 y 0.90 exactitud moderada, y más de 0.90 exactitud alta. También de manera algo arbitraria, otros investigadores han señalado valores de referencia ligeramente más exigentes aunque, en general, compatibles con los anteriores; así, por ejemplo, Fan et al. (2006) sugieren que el área de la curva de ROC debe ser mayor de 0.75 para tener significación clínica y mayor de 0.97 para un valor clínico alto.

El área bajo la curva ROC es el estadístico que más se recomienda en los últimos años para el análisis de la eficacia diagnóstica (Rossi y Sloore, 2005). Además, permite detectar puntos de corte que tengan un buen equilibrio entre sensibilidad y especificidad.

Todos estos estadísticos de eficacia o rendimiento diagnóstico fueron calculados para realizar las siguientes discriminaciones a partir de los puntos de corte propuestos por los manuales del MCMI-II y del BDI-II (p. ej., puntuaciones TB 75 y 85 en el caso del MCMI-II) u otros sugeridos por los propios estadísticos de rendimiento diagnóstico:

- 1) Entre pacientes con un diagnóstico de trastorno depresivo mayor y pacientes con otro tipo de diagnóstico según el DSM-IV.
- 2) Entre pacientes con cualquier diagnóstico del conjunto de trastornos depresivos (trastorno depresivo mayor, distímico, depresivo no especificado, adaptativo con estado de ánimo depresivo, adaptativo mixto y duelo) y pacientes con un diagnóstico no depresivo.
- 3) Entre pacientes con un diagnóstico de trastorno depresivo mayor y pacientes con un diagnóstico no depresivo (excluyendo, por tanto, entre estos últimos pacientes, aquéllos con un diagnóstico de otros trastornos depresivos: trastorno distímico, trastorno depresivo no especificado, trastorno adaptativo con estado de ánimo depresivo, trastorno adaptativo mixto y duelo, o con un diagnóstico secundario de trastorno depresivo).

Las curvas ROC y los diferentes índices de rendimiento diagnóstico fueron calculados mediante el paquete estadístico IBM SPSS, versión 19. Para comparar la validez diagnóstica de las diferentes medidas, escalas e instrumentos (puntuaciones directas y TB de las escalas CC y D del MCMI-II y puntuaciones directas del BDI-II), se compararon las curvas ROC correspondientes mediante el programa estadístico EPIDAT,

versión 3.1 (Xunta de Galicia y Organización Panamericana de la Salud, Santiago de Compostela, España; <http://dxsp.sergas.es>) y se compararon dos a dos las áreas de las curvas ROC mediante el estadístico z del programa estadístico MedCalc para Windows, versión 12.4.0 (MedCalc Software, Mariakerke, Bélgica; <http://www.medcalc.org>).

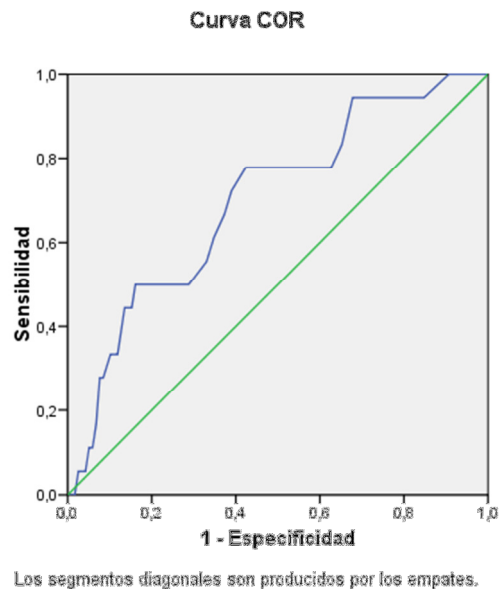
### 6.3. Resultados

#### **Eficacia diagnóstica del MCMI-II y del BDI-II para discriminar entre pacientes con un diagnóstico de trastorno depresivo mayor y pacientes con otro tipo de diagnóstico según el DSM-IV**

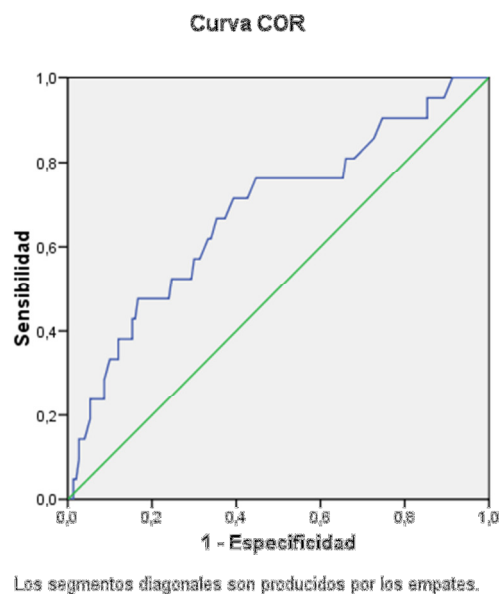
Las curvas ROC realizadas sobre cada una de las puntuaciones directas y de tasa base de las escalas CC y D del MCMI-II, así como sobre la puntuación total del BDI-II, en relación con el diagnóstico de trastorno depresivo mayor realizado por los clínicos, se recogen en las Figuras 6.2, 6.3, 6.4, 6.5 y 6.6.

Se puede observar a través de un análisis visual de las Figuras 6.2 a 6.5 que las curvas ROC de la escala D del MCMI-II para el diagnóstico del trastorno depresivo mayor, tanto si se tienen en cuenta sus puntuaciones brutas como sus puntuaciones TB, se separaban moderadamente de los segmentos diagonales producidos por los empates entre sensibilidad y tasa de falsos positivos que reflejan un diagnóstico basado en el azar (representado en las gráficas mediante la línea diagonal verde). En cambio, las puntuaciones brutas y tasa base de la escala de CC se separaban tan solo ligeramente de los segmentos diagonales.

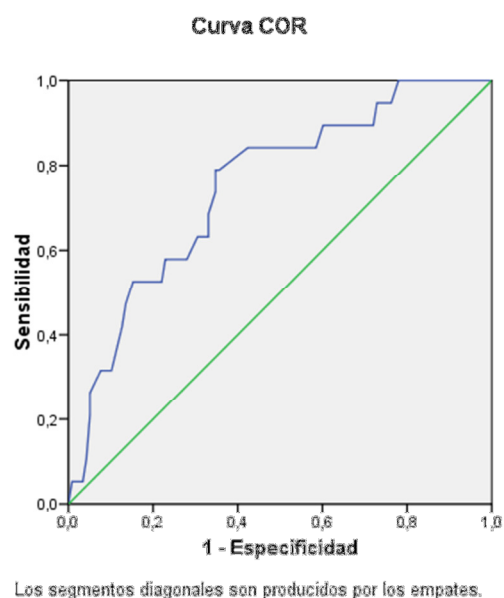
**Figura 6.2.** Curva ROC de las puntuaciones directas de la escala de Depresión Mayor (CC) del MCMI-II con respecto al diagnóstico de trastorno depresivo mayor



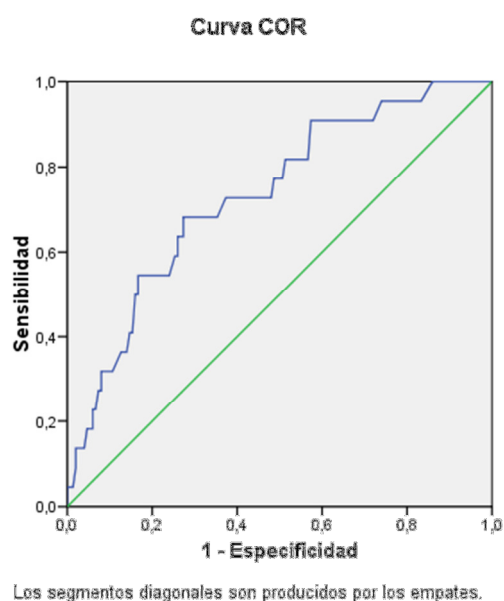
**Figura 6.3.** Curva ROC de las puntuaciones tasa base (TB) de la escala de Depresión Mayor (CC) del MCMI-II con respecto al diagnóstico trastorno depresivo mayor



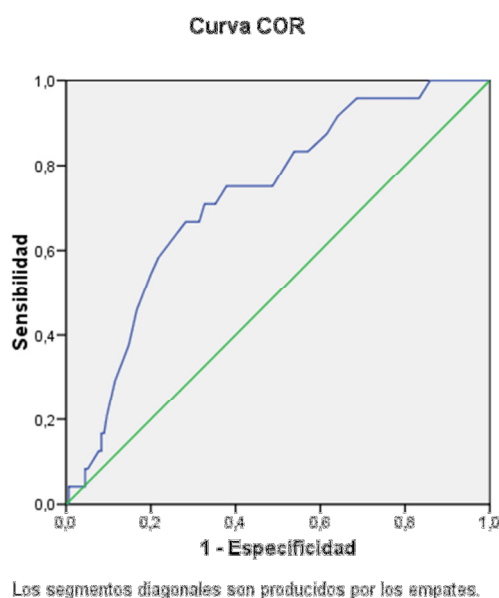
**Figura 6.4.** Curva ROC de las puntuaciones directas de la escala de Distimia (D) del MCMI-II con respecto al diagnóstico de trastorno depresivo mayor



**Figura 6.5.** Curva ROC de las puntuaciones tasa base (TB) de la escala de Distimia (D) del MCMI-II con respecto al diagnóstico de trastorno depresivo mayor



**Figura 6.6.** Curva ROC de las puntuaciones del BDI-II con respecto al diagnóstico de trastorno depresivo mayor



Como se puede observar en la Tabla 6.9, las áreas bajo la curva ROC de las puntuaciones brutas y tasa base de la escala D del MCMI-II y de las puntuaciones tasa base de la escala CC del MCMI-II fueron estadísticamente significativas con  $p < 0.05$ , es decir, fueron significativamente mayores que el área mínima de 0.5 que representa un diagnóstico basado en el azar. Es más, las áreas bajo la curva ROC de las puntuaciones directas y TB de la escala de D y de las puntuaciones tasa base de la escala CC, superaron al menos el criterio mínimo propuesto por algunos autores para un instrumento diagnóstico ( $> 0.70$ , según Swets, 1988), aunque no el criterio más estricto recomendado por otros ( $> 0.75$ , según Fan et al., 2006). Sin embargo, el área bajo la curva ROC de las puntuaciones directas de la escala CC del MCMI-II no sólo no fue estadísticamente significativa sino que tampoco alcanzó los criterios mínimos recomendados para un instrumento diagnóstico ( $> 0.70$ , según Swets, 1988, o  $> 0.75$ , según Fan et al., 2006).



**Tabla 6.9.** Áreas bajo la curva ROC en relación con el diagnóstico de trastorno depresivo mayor de las puntuaciones directas y tasa base de las escalas de Depresión Mayor (CC) y Distimia (D) del MCMI-II y de las puntuaciones directas del BDI-II

Puntuación — Escala	Área bajo ROC para diagnóstico depresión mayor	Intervalo de confianza asintótico del área al 95%	
		Límite inferior	Límite superior
PD — CC	0.699	0.571	0.826
TB — CC	0.710*	0.583	0.836
PD — D	0.742*	0.624	0.859
TB — D	0.738*	0.628	0.849
BDI-II	0.771*	0.671	0.871

*Nota.* \* Estadísticamente significativa con  $p < 0.001$ . PD — CC = Puntuación directa de la escala de Depresión Mayor del MCMI-II; TB — CC = Puntuación tasa base de la escala de Depresión Mayor del MCMI-II; PD — D = Puntuación directa de la escala de Distimia del MCMI-II; TB — D = Puntuación tasa base de la escala de Distimia del MCMI-II; BDI-II = Puntuación directa del Inventario de Depresión de Beck-II.

Respecto al BDI-II, el análisis visual de la Figura 6.6 que presenta su curva ROC para el diagnóstico del trastorno depresivo mayor también sugería que dicha curva se separaba moderadamente de la línea diagonal que refleja un diagnóstico basado en el azar (representado en la Figura 6.6 mediante la línea diagonal verde). De hecho, también su área bajo la curva ROC fue estadísticamente significativa con  $p < 0.05$ , es decir, fue significativamente mayor que el área mínima de 0.5 que representa un diagnóstico basado en el azar (véase la Tabla 6.9), alcanzando no sólo el criterio mínimo recomendado para un instrumento diagnóstico por algunos autores ( $> 0.70$ , según Swets,

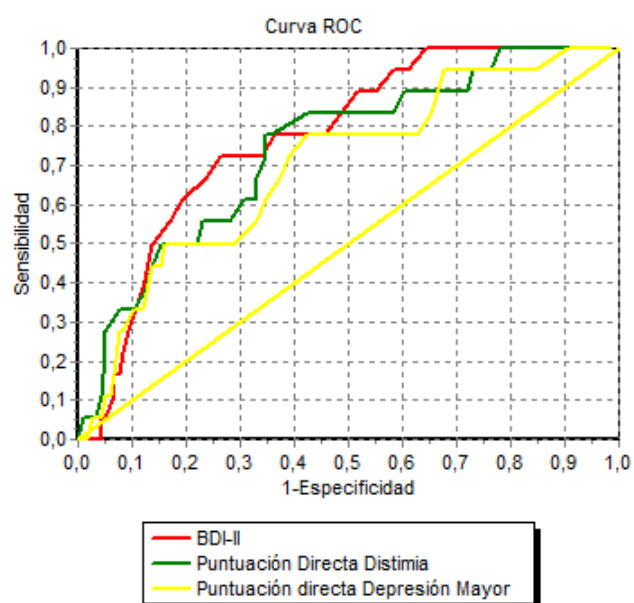
1988), sino también el criterio más estricto recomendado por otros ( $> 0.75$ , según Fan et al., 2006).

Sin embargo, cuando se compararon, respecto al diagnóstico de trastorno depresivo mayor, las curvas ROC de las escalas CC y D del MCMI-II y del BDI-II, no se encontraron diferencias estadísticamente significativas entre sus respectivas áreas bajo las curvas ROC, es decir, entre los índices de validez diagnóstica de esas tres medidas de depresión.

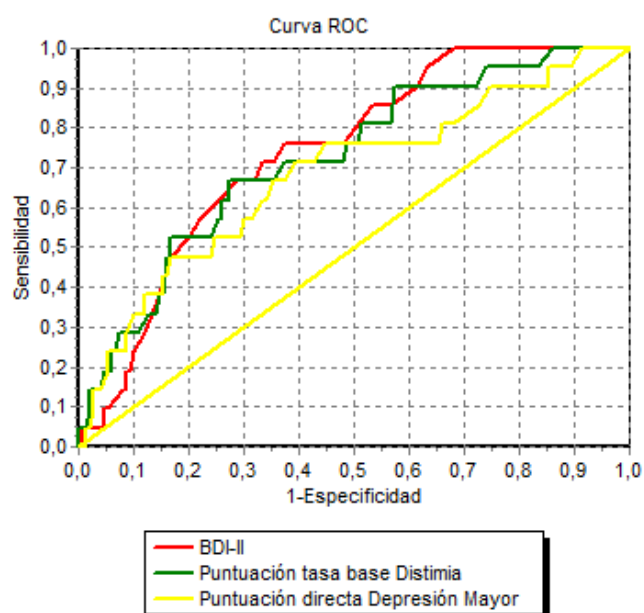
Efectivamente, como puede verse en la Figura 6.7, las curvas ROC de las puntuaciones directas de las escalas CC y D del MCMI-II y del BDI-II en relación al diagnóstico de trastorno depresivo mayor fueron muy similares, por lo que parece lógico que el área bajo la curva ROC de las puntuaciones directas de la escala D no difiriera significativamente de las áreas bajo la curva ROC de las puntuaciones directas de la escala CC ( $z = 1.31$ ,  $p = 0.188$ ) o del BDI-II ( $z = 0.51$ ,  $p = 0.607$ ), ni que tampoco difirieran entre sí de forma significativa las áreas bajo la curva ROC de las puntuaciones directas de la escala CC y del BDI-II ( $z = 1.72$ ,  $p = 0.08$ ) (véase también el Anexo 7).

De forma similar, y tal y como puede observarse en la Figura 6.8, las curvas ROC de las puntuaciones tasa base de las escalas CC y D del MCMI-II y de las puntuaciones directas del BDI-II en relación al diagnóstico de trastorno depresivo mayor fueron muy similares, y, de nuevo, el área bajo la curva ROC de las puntuaciones tasa base de la escala D no difirió significativamente de las áreas bajo la curva ROC de las puntuaciones tasa base de la escala CC ( $z = 1.32$ ,  $p = 0.184$ ) o de las puntuaciones directas del BDI-II ( $z = 0.32$ ,  $p = 0.742$ ), ni tampoco difirieron entre sí de forma significativa las áreas bajo la curva ROC de las puntuaciones tasa base de la escala CC y de las puntuaciones directas del BDI-II ( $z = 1.07$ ,  $p = 0.281$ ) (véase también el Anexo 8).

**Figura 6.7.** Curvas ROC de las puntuaciones directas de las escalas de Distimia (D) y Depresión Mayor (CC) del MCMI-II y del BDI-II con respecto al diagnóstico de trastorno depresivo mayor



**Figura 6.8.** Curvas ROC de las puntuaciones tasa base de las escalas de Distimia (D) y Depresión Mayor (CC) del MCMI-II y de las puntuaciones directas del BDI-II con respecto al diagnóstico de trastorno depresivo mayor



### Puntos de corte del MCMI-II y del BDI-II para discriminar entre pacientes con un diagnóstico de trastorno depresivo mayor y pacientes con otro tipo de diagnóstico según el DSM-IV

En las Tablas 6.10, 6.11 y 6.12 se recogen las coordenadas de las curvas ROC de las escalas CC y D del MCMI-II y del BDI-II respecto al diagnóstico del trastorno depresivo mayor. Estas coordenadas representan la sensibilidad y especificidad específicas de cada una de las puntuaciones directas y puntuaciones tasa base (TB) de dichos instrumentos. Así, en la Tabla 6.10 se puede observar que, para el trastorno depresivo mayor, se alcanzaron sensibilidades superiores a 0.70 en la escala CC del MCMI-II con puntuaciones directas inferiores a 19 y puntuaciones tasa base inferiores a 52, mientras que especificidades superiores a 0.70 suponían en la escala CC puntuaciones directas superiores a 22 y puntuaciones tasa base superiores a 60.

**Tabla 6.10.** Coordenadas de las curvas ROC de las puntuaciones directas y tasa base de la escala de Depresión Mayor (CC) del MCMI-II con respecto al diagnóstico de trastorno depresivo mayor

Positivo si la puntuación directa de CC es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad	Positivo si la puntuación tasa base de CC es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad
-1.00	1.000	1.000	-1.00	1.000	1.000
0.50	1.000	0.983	1.50	1.000	0.980
1.50	1.000	0.932	3.50	1.000	0.973
2.50	1.000	0.907	4.50	1.000	0.967
3.50	0.944	0.847	5.50	1.000	0.940
4.50	0.944	0.814	7.00	1.000	0.933
5.50	0.944	0.771	8.50	1.000	0.920
6.50	0.944	0.746	9.50	1.000	0.913
7.50	0.944	0.720	11.50	0.952	0.893
8.50	0.944	0.695	13.50	0.952	0.860
9.50	0.944	0.678	14.50	0.952	0.853
10.50	0.833	0.653	15.50	0.905	0.853

Positivo si la puntuación directa de CC es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad	Positivo si la puntuación tasa base de CC es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad
11.50	0.778	0.627	16.50	0.905	0.840
13.00	0.778	0.610	17.50	0.905	0.833
14.50	0.778	0.551	18.50	0.905	0.827
15.50	0.778	0.508	20.50	0.905	0.807
16.50	0.778	0.449	23.00	0.905	0.760
17.50	0.778	0.424	25.00	0.905	0.747
18.50	0.722	0.390	28.00	0.857	0.727
19.50	0.667	0.373	30.50	0.810	0.680
20.50	0.611	0.347	32.50	0.810	0.667
21.50	0.556	0.331	35.00	0.810	0.660
22.50	0.500	0.288	36.50	0.762	0.653
23.50	0.500	0.254	37.50	0.762	0.647
24.50	0.500	0.229	38.50	0.762	0.627
25.50	0.500	0.220	39.50	0.762	0.613
26.50	0.500	0.203	40.50	0.762	0.587
27.50	0.500	0.195	41.50	0.762	0.580
28.50	0.500	0.186	42.50	0.762	0.540
29.50	0.500	0.178	43.50	0.762	0.527
30.50	0.500	0.161	44.50	0.762	0.520
31.50	0.444	0.153	45.50	0.762	0.480
32.50	0.444	0.136	46.50	0.762	0.473
33.50	0.333	0.119	47.50	0.762	0.447
34.50	0.333	0.102	48.50	0.714	0.427
35.50	0.278	0.085	50.00	0.714	0.420
36.50	0.278	0.076	51.50	0.714	0.393
37.50	0.167	0.068	52.50	0.667	0.373
38.50	0.111	0.059	53.50	0.667	0.367
39.50	0.111	0.051	54.50	0.667	0.360
40.50	0.056	0.042	55.50	0.667	0.353
43.00	0.056	0.025	56.50	0.619	0.340
46.50	0.000	0.017	57.50	0.619	0.333
49.50	0.000	0.008	58.50	0.571	0.313
52.00	0.000	0.000	59.50	0.571	0.300
			60.50	0.524	0.293
			61.50	0.524	0.287
			62.50	0.524	0.280
			63.50	0.524	0.260
			64.50	0.524	0.253
			65.50	0.524	0.247
			67.00	0.476	0.240
			68.50	0.476	0.233
			69.50	0.476	0.227
			70.50	0.476	0.220

Positivo si la puntuación directa de CC es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad	Positivo si la puntuación tasa base de CC es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad
			71.50	0.476	0.200
			72.50	0.476	0.180
			74.00	0.476	0.167
			75.50	0.429	0.160
			77.50	0.429	0.153
			79.50	0.381	0.153
			80.50	0.381	0.147
			81.50	0.381	0.133
			83.00	0.381	0.127
			84.50	0.381	0.120
			86.00	0.333	0.120
			88.00	0.333	0.100
			89.50	0.286	0.087
			90.50	0.238	0.087
			91.50	0.238	0.080
			93.00	0.238	0.067
			94.50	0.238	0.053
			95.50	0.190	0.053
			96.50	0.143	0.040
			100.00	0.143	0.027
			104.00	0.095	0.027
			105.50	0.048	0.020
			107.50	0.048	0.013
			109.50	0.000	0.013
			111.00	0.000	0.000

*Nota.* <sup>a</sup>El punto de corte menor es la mínima puntuación observada menos 1, mientras que el punto de corte mayor es la máxima puntuación observada más 1. Todos los demás puntos de corte son la media de dos puntuaciones observadas ordenadas y consecutivas. CC = Escala de Depresión Mayor del MCMI-II.

En la misma dirección, en la Tabla 6.11 se recogen las coordenadas de las curvas ROC de las puntuaciones directas y de TB de la escala de D del MCMI-II con respecto al diagnóstico de trastorno depresivo mayor, las cuales indican que sensibilidades superiores a 0.70 se alcanzaban con puntuaciones directas y de TB inferiores a 28 y 66, respectivamente, mientras que especificidades superiores a 0.70 suponían puntuaciones directas y puntuaciones de TB superiores a 31 y 70, respectivamente.

**Tabla 6.11.** Coordenadas de las curvas ROC de las puntuaciones directas y de tasa base de la escala de Distimia del MCMI-II con respecto al diagnóstico de trastorno depresivo mayor

<b>Positivo si la puntuación directa de D es mayor o igual que<sup>a</sup></b>	<b>Sensibilidad</b>	<b>1 - especificidad</b>	<b>Positivo si la puntuación tasa base de D es mayor o igual que<sup>a</sup></b>	<b>Sensibilidad</b>	<b>1 - especificidad</b>
-1.00	1.000	1.000	-1.00	1.000	1.000
0.50	1.000	0.992	4.00	1.000	0.993
1.50	1.000	0.983	10.00	1.000	0.987
2.50	1.000	0.958	12.50	1.000	0.980
3.50	1.000	0.941	14.50	1.000	0.973
4.50	1.000	0.898	16.50	1.000	0.967
5.50	1.000	0.864	18.00	1.000	0.960
6.50	1.000	0.822	19.50	1.000	0.947
7.50	1.000	0.780	21.00	1.000	0.927
8.50	0.947	0.763	22.50	1.000	0.913
9.50	0.947	0.746	23.50	1.000	0.900
10.50	0.947	0.729	24.50	1.000	0.880
11.50	0.895	0.720	26.00	1.000	0.873
12.50	0.895	0.703	27.50	1.000	0.867
13.50	0.895	0.669	28.50	1.000	0.860
14.50	0.895	0.653	29.50	0.955	0.833
15.50	0.895	0.619	30.50	0.955	0.813
16.50	0.895	0.610	32.00	0.955	0.807
17.50	0.895	0.602	33.50	0.955	0.773
18.50	0.842	0.585	34.50	0.955	0.760
19.50	0.842	0.559	35.50	0.955	0.747
20.50	0.842	0.508	37.00	0.955	0.740
21.50	0.842	0.475	38.50	0.909	0.720
22.50	0.842	0.449	39.50	0.909	0.713
23.50	0.842	0.441	40.50	0.909	0.707
24.50	0.842	0.424	41.50	0.909	0.673
25.50	0.789	0.356	42.50	0.909	0.647
26.50	0.789	0.347	43.50	0.909	0.620
27.50	0.737	0.347	44.50	0.909	0.593
28.50	0.684	0.331	45.50	0.909	0.587
29.50	0.632	0.331	46.50	0.909	0.573
30.50	0.632	0.305	47.50	0.818	0.567
31.50	0.579	0.280	48.50	0.818	0.560
32.50	0.579	0.263	49.50	0.818	0.547
33.50	0.579	0.254	50.50	0.818	0.527
34.50	0.579	0.237	51.50	0.818	0.513
35.50	0.579	0.229	52.50	0.773	0.507
37.00	0.526	0.220	54.00	0.773	0.487

Positivo si la puntuación directa de D es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad	Positivo si la puntuación tasa base de D es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad
39.00	0.526	0.203	56.00	0.727	0.480
40.50	0.526	0.195	57.50	0.727	0.460
41.50	0.526	0.186	58.50	0.727	0.440
42.50	0.526	0.178	59.50	0.727	0.433
43.50	0.526	0.153	61.00	0.727	0.427
44.50	0.474	0.136	62.50	0.727	0.420
45.50	0.421	0.127	63.50	0.727	0.400
47.00	0.316	0.102	64.50	0.727	0.373
49.00	0.316	0.076	66.00	0.682	0.353
50.50	0.263	0.051	68.00	0.682	0.320
51.50	0.211	0.051	69.50	0.682	0.307
53.00	0.105	0.042	70.50	0.682	0.293
54.50	0.053	0.034	71.50	0.682	0.287
55.50	0.053	0.017	72.50	0.682	0.273
58.00	0.053	0.008	73.50	0.636	0.273
61.00	0.000	0.000	74.50	0.636	0.267
			75.50	0.636	0.260
			76.50	0.591	0.260
			77.50	0.591	0.253
			78.50	0.545	0.240
			80.00	0.545	0.233
			81.50	0.545	0.227
			82.50	0.545	0.220
			83.50	0.545	0.200
			84.50	0.545	0.180
			85.50	0.545	0.167
			86.50	0.500	0.167
			87.50	0.500	0.160
			88.50	0.409	0.153
			90.00	0.409	0.147
			91.50	0.364	0.140
			92.50	0.364	0.133
			93.50	0.364	0.127
			94.50	0.318	0.107
			95.50	0.318	0.093
			96.50	0.318	0.080
			97.50	0.273	0.080
			98.50	0.273	0.073
			100.50	0.227	0.067
			103.00	0.227	0.060
			104.50	0.182	0.060
			105.50	0.182	0.047
			106.50	0.136	0.040



Positivo si la puntuación directa de D es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad	Positivo si la puntuación tasa base de D es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad
			107.50	0.136	0.020
			108.50	0.091	0.020
			109.50	0.045	0.013
			111.50	0.045	0.007
			118.50	0.045	0.000
			125.00	0.000	0.000

*Nota.* <sup>a</sup>El punto de corte menor es la mínima puntuación observada menos 1, mientras que el punto de corte mayor es la máxima puntuación observada más 1. Todos los demás puntos de corte son la media de dos puntuaciones observadas ordenadas y consecutivas. D = Escala de Distimia del MCMI-II.

Finalmente, en la Tabla 6.12 se recogen las coordenadas de las curvas ROC de las puntuaciones del BDI-II con respecto al diagnóstico de trastorno depresivo mayor, las cuales indican que sensibilidades superiores a 0.70 se alcanzaban con puntuaciones del BDI-II inferiores a 23, mientras que especificidades superiores a 0.70 suponían puntuaciones superiores a 25.

**Tabla 6.12.** Coordenadas de las curvas ROC de las puntuaciones del BDI-II con respecto al diagnóstico de trastorno depresivo mayor

Positivo si la puntuación total del BDI-II es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad
-1.00	1.000	1.000
0.50	1.000	0.987
1.50	1.000	0.981
2.50	1.000	0.974
3.50	1.000	0.962
4.50	1.000	0.942
5.50	1.000	0.910
6.50	1.000	0.891
7.50	1.000	0.859
8.50	0.958	0.833
9.50	0.958	0.782
10.50	0.958	0.769
11.50	0.958	0.731
12.50	0.958	0.686
13.50	0.917	0.641

Positivo si la puntuación total del BDI-II es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad
14.50	0.875	0.615
15.50	0.833	0.571
16.50	0.833	0.538
17.50	0.750	0.487
18.50	0.750	0.442
19.50	0.750	0.397
20.50	0.750	0.378
21.50	0.708	0.353
22.50	0.708	0.327
23.50	0.667	0.314
24.50	0.667	0.308
25.50	0.667	0.282
26.50	0.625	0.250
27.50	0.583	0.218
28.50	0.542	0.199
29.50	0.458	0.167
30.50	0.375	0.147
31.50	0.292	0.115
32.50	0.208	0.096
33.50	0.167	0.090
34.50	0.167	0.083
35.50	0.125	0.083
36.50	0.125	0.077
37.50	0.083	0.051
38.50	0.083	0.045
39.50	0.042	0.045
40.50	0.042	0.038
41.50	0.042	0.032
42.50	0.042	0.026
44.00	0.042	0.006
47.50	0.000	0.006
51.00	0.000	0.000

*Nota.* <sup>a</sup>El punto de corte menor es la mínima puntuación observada menos 1, mientras que el punto de corte mayor es la máxima puntuación observada más 1. Todos los demás puntos de corte son la media de dos puntuaciones observadas ordenadas y consecutivas.

En las Tablas 6.13, 6.14 y 6.15, además de la sensibilidad y la especificidad, se recogen un mayor número de índices de validez diagnóstica (índice de Youden, poder predictivo positivo, poder predictivo negativo, eficacia global y kappa) en relación con el diagnóstico de trastorno depresivo mayor para las puntuaciones de las escalas CC y D

del MCMI-II y del BDI-II que han sido propuestas en los manuales correspondientes como puntos de corte para el diagnóstico de trastorno depresivo mayor (puntuaciones TB 75 y TB 85 en el MCMI-II) o para la delimitación de diferentes niveles de gravedad de la sintomatología depresiva (puntuaciones en el BDI-II igual a 14, 20 y 28 para delimitar niveles de depresión leve, moderada y grave, respectivamente). Así mismo, se presentan esos mismos índices de validez diagnóstica para algunas puntuaciones que se encuentran cercanas a las anteriores.

Como se puede comprobar en la Tabla 6.13, los puntos de corte propuestos por los autores del MCMI-II (TB 75 y TB 85) son los que en la presente investigación tuvieron, en general, los valores más elevados de concordancia con el diagnóstico de referencia de trastorno depresivo mayor ( $\kappa$ ), aunque, sin embargo, dichos valores  $\kappa$  estaban por debajo de los valores mínimos de referencia para una concordancia diagnóstica moderada ( $\kappa > 0.40$ ) y se encontraban en el rango de valores más bajos para una concordancia diagnóstica débil ( $\kappa > 0.20$ ). Así, en el caso de la escala CC, el valor más elevado se encontró en las puntuaciones TB 75 y TB 85, ambos con  $\kappa = 0.24$ , y la puntuación con mayor índice de Youden fue la puntuación TB 75 (Youden = 30.95). Estos puntos de corte tuvieron una alta especificidad, por encima de 90%, pero una débil sensibilidad: 47.62% en el caso de la puntuación TB 75 y 38.1% en la puntuación TB 85, ambos por debajo del valor mínimo de referencia para los índices de sensibilidad ( $> 70\%$ ). Sin embargo, el índice de eficacia global de la TB 75 fue 78.95% y de la TB 85 fue 81.87%. La puntuación TB 60 mostraba un mayor equilibrio entre especificidad (70%) y sensibilidad (57.14%) y esta última continuaba por debajo del valor de referencia ( $> 70\%$ ), y con menores valores en otros índices de eficacia:  $\kappa$  (0.16), índice de Youden (27.14) y eficacia global (68.42%).

La puntuación directa en la escala CC con mayor kappa respecto al diagnóstico de trastorno depresivo realizado por el clínico fue la puntuación de 32 ( $\text{kappa} = 0.25$ ), con una especificidad de 84.75% y una sensibilidad de 44.44% (véase la Tabla 6.13). Sin embargo, la puntuación directa de 18 mostró el mayor índice de Youden (35.41), con un mayor equilibrio entre sensibilidad y especificidad (77.78% y 57.63%, respectivamente). Por otro lado, la puntuación directa de 18 en la escala CC tuvo una eficacia global de 60.29%, mientras que en el caso de la puntuación directa 32 dicho índice fue de 79.41%. En general, se encontró para las puntuaciones directas de referencia en la escala CC un poder predictivo negativo muy alto (superior a 90%) y un poder predictivo positivo muy débil (inferior a 30%).

Por su parte, en el caso de la escala de D (véase la Tabla 6.14), la puntuación TB 85 obtuvo un  $\text{kappa} = 0.27$  y la puntuación TB 75 obtuvo un  $\text{kappa} = 0.23$ , los cuales se sitúan en el rango inferior de los valores de referencia para una concordancia diagnóstica débil ( $\text{kappa} > 0.20$ ; Altman, 1991). La puntuación TB 85 tuvo una sensibilidad de 54.55% y una especificidad de 82%, mientras que la puntuación TB 75 mostró una sensibilidad mayor (63.64%) y una menor especificidad (73.33%), por lo que ambas puntuaciones obtuvieron un similar índice de Youden (36.55 y 36.97, respectivamente). La puntuación TB 85 obtuvo un mayor índice de eficacia global (78.49, frente al 72.09 de la puntuación TB 75). La puntuación de 23 en la escala D fue la puntuación directa que presentó un mayor índice de Youden (39.29), con una alta sensibilidad de 84.21% y una especificidad de 55.08%, aunque con un índice de concordancia menor ( $\text{kappa} = 0.19$ ) y una eficacia global de 59.12%. Al igual que ocurría en la escala CC, en la escala D los distintos puntos de corte también mostraron un poder predictivo negativo muy alto (superior a 90%) y un poder predictivo positivo débil (inferior a 30%).

En el caso del BDI-II (véase la Tabla 6.15), la puntuación de 28 presentó el mayor índice kappa (0.26) y un índice de Youden de 36.54, el tercero más alto después de la puntuación de 24 y 27. El punto de corte 28 obtuvo una especificidad de 78.21% y una sensibilidad de 58.33%. Otras puntuaciones del BDI-II con niveles más adecuados de sensibilidad, es decir, por encima de 70%, como, por ejemplo, la puntuación 21 (sensibilidad = 75%), tenían índices kappa inferiores (0.19), aunque un índice de Youden también superior a 34. La puntuación de 14 en el BDI-II presentó en este estudio valores muy elevados de sensibilidad (91.67%), pero a costa de un nivel de especificidad bajo (35.9%). También para las puntuaciones del BDI-II analizadas, se encontraron índices muy elevados de poder predictivo negativo (en general > 90%), pero muy bajos de poder predictivo positivo (< 30%).

**Tabla 6.13.** Índices de validez diagnóstica de distintas puntuaciones directas (PD) o tasa base (TB) de la escala de Depresión Mayor (CC) del MCMI-II con respecto al diagnóstico del trastorno depresivo mayor

Punto de corte – Escala	Sensibilidad	Especificidad	Índice de Youden	Poder predictivo positivo	Poder predictivo negativo	Tasa de falsos positivos	Tasa de falsos negativos	Kappa	Eficacia
<b>TB 75</b>	<b>47.62</b>	<b>83.33</b>	<b>30.95</b>	<b>28.57</b>	<b>91.91</b>	<b>16.67</b>	<b>52.38</b>	<b>0.24</b>	<b>78.95</b>
<b>TB 85</b>	<b>38.1</b>	<b>88</b>	<b>26.1</b>	<b>30.77</b>	<b>91.03</b>	<b>12</b>	<b>61.9</b>	<b>0.24</b>	<b>81.87</b>
TB 66	52.38	75.33	27.71	22.92	91.87	24.67	47.62	0.18	72.51
TB 67	47.62	76	23.62	21.74	91.2	24	52.38	0.16	72.51
TB 68	47.62	76	23.62	21.74	91.2	24	52.38	0.16	72.51
TB 62	52.38	71.33	23.71	20.37	91.45	28.67	47.62	0.14	69.01
TB 61	52.38	70.67	23.05	20	91.38	29.33	47.62	0.14	68.42
<b>TB 60</b>	<b>57.14</b>	<b>70</b>	<b>27.14</b>	<b>21.05</b>	<b>92.11</b>	<b>30</b>	<b>42.86</b>	<b>0.16</b>	<b>68.42</b>
PD 7	94.44	25.42	19.86	16.19	96.77	74.58	5.56	0.07	34.56
PD 9	94.44	30.51	24.95	17.17	97.3	69.49	5.56	0.09	38.97
PD 8	94.44	27.97	22.41	16.67	97.06	72.03	5.56	0.08	36.76
PD 10	94.44	32.2	26.64	17.53	97.44	67.8	5.56	0.09	40.44
PD 11	83.33	34.75	18.08	16.3	93.18	65.25	16.67	0.07	41.18
PD 12	77.78	37.29	15.07	15.91	91.67	62.71	22.22	0.06	42.65
PD 13	77.78	38.98	16.76	16.28	92	61.02	22.22	0.06	44.12
PD 14	77.78	38.98	16.76	16.28	92	61.02	22.22	0.06	44.12
PD 15	77.78	44.92	22.7	17.72	92.98	55.08	22.22	0.09	49.26
PD 16	77.78	49.15	26.93	18.92	93.55	50.85	22.22	0.12	52.94
PD 17	77.78	55.08	32.86	20.9	94.2	44.92	22.22	0.15	58.09
<b>PD 18</b>	<b>77.78</b>	<b>57.63</b>	<b>35.41</b>	<b>21.88</b>	<b>94.44</b>	<b>42.37</b>	<b>22.22</b>	<b>0.17</b>	<b>60.29</b>
PD 19	72.22	61.02	33.24	22.03	93.51	38.98	27.78	0.17	62.5
PD 20	66.67	62.71	29.38	21.43	92.5	37.29	33.33	0.16	63.24
PD 23	50	71.19	21.19	20.93	90.32	28.81	50	0.13	68.38
PD 28	50	80.51	30.51	28.13	91.35	19.49	50	0.23	76.47
<b>PD 32</b>	<b>44.44</b>	<b>84.75</b>	<b>29.19</b>	<b>30.77</b>	<b>90.91</b>	<b>15.25</b>	<b>55.56</b>	<b>0.25</b>	<b>79.41</b>
PD 35	33.33	89.83	23.16	33.33	89.83	10.17	66.67	0.23	82.35

**Tabla 6.14.** Índices de validez diagnóstica de distintas puntuaciones directas (PD) o tasa base (TB) de la escala de Distimia (D) del MCMI-II con respecto al diagnóstico del trastorno depresivo mayor

Punto de corte – Escala	Sensibilidad	Especificidad	Índice de Youden	Poder predictivo positivo	Poder predictivo negativo	Tasa de falsos positivos	Tasa de falsos negativos	Kappa	Eficacia
<b>TB 85</b>	<b>54.55</b>	<b>82</b>	<b>36.55</b>	<b>30.77</b>	<b>92.48</b>	<b>18</b>	<b>45.45</b>	<b>0.27</b>	<b>78.49</b>
<b>TB 75</b>	<b>63.64</b>	<b>73.33</b>	<b>36.97</b>	<b>25.93</b>	<b>93.22</b>	<b>26.67</b>	<b>36.36</b>	<b>0.23</b>	<b>72.09</b>
TB 66	68.18	64.67	32.85	22.06	93.27	35.33	31.82	0.17	65.12
TB 67	68.18	64.67	32.85	22.06	93.27	35.33	31.82	0.17	65.12
<b>TB 68</b>	<b>68.18</b>	<b>68</b>	<b>36.18</b>	<b>23.81</b>	<b>93.58</b>	<b>32</b>	<b>31.82</b>	<b>0.2</b>	<b>68.02</b>
TB 60	72.73	56.67	29.4	19.75	93.41	43.33	27.27	0.14	58.72
TB 61	72.73	57.33	30.06	20	93.48	42.67	27.27	0.14	59.3
TB 62	72.73	57.33	30.06	20	93.48	42.67	27.27	0.14	59.3
PD 8	100	22.03	22.03	17.12	100	77.97	0	0.07	32.85
PD 9	94.74	23.73	18.47	16.67	96.55	76.27	5.26	0.06	33.58
PD 10	94.74	25.42	20.16	16.98	96.77	74.58	5.26	0.07	35.04
PD 11	94.74	27.12	21.86	17.31	96.97	72.88	5.26	0.08	36.5
PD 15	89.47	34.75	24.22	18.09	95.35	65.25	10.53	0.09	42.34
PD 18	89.47	39.83	29.3	19.32	95.92	60.17	10.53	0.12	46.72
PD 20	84.21	44.07	28.28	19.51	94.55	55.93	15.79	0.12	49.64
<b>PD 23</b>	<b>84.21</b>	<b>55.08</b>	<b>39.29</b>	<b>23.19</b>	<b>95.59</b>	<b>44.92</b>	<b>15.79</b>	<b>0.19</b>	<b>59.12</b>
PD 28	73.68	65.25	38.93	25.45	93.9	34.75	26.32	0.22	66.42
PD 32	57.89	72.03	29.92	25	91.4	27.97	42.11	0.19	70.07
PB 35	57.89	76.27	34.16	28.21	91.84	23.73	42.11	0.24	73.72

**Tabla 6.15.** Índices de validez diagnóstica de distintas puntuaciones del BDI-II con respecto al diagnóstico de trastorno depresivo mayor

Punto de corte — Escala	Sensibilidad	Especificidad	Índice de Youden	Poder predictivo positivo	Poder predictivo negativo	Tasa de falsos positivos	Tasa de falsos negativos	Kappa	Eficacia
<b>BDI-II 14</b>	<b>91.67</b>	<b>35.9</b>	<b>27.57</b>	<b>18.03</b>	<b>96.55</b>	<b>64.1</b>	<b>8.33</b>	<b>0.1</b>	<b>43.33</b>
BDI-II 15	87.5	38.46	25.96	17.95	95.24	61.54	12.5	0.1	45
BDI-II 16	83.33	42.95	26.28	18.35	94.37	57.05	16.67	0.11	48.33
BDI-II 17	83.33	46.15	29.48	19.23	94.74	53.85	16.67	0.12	51.11
BDI-II 18	75	51.28	26.28	19.15	93.02	48.72	25	0.12	54.44
BDI-II 19	75	55.77	30.77	20.69	93.55	44.23	25	0.15	58.33
BDI-II 20	75	60.26	35.26	22.5	94	39.74	25	0.18	62.22
<b>BDI-II 21</b>	<b>75</b>	<b>62.18</b>	<b>37.18</b>	<b>23.38</b>	<b>94.17</b>	<b>37.82</b>	<b>25</b>	<b>0.19</b>	<b>63.89</b>
BDI-II 22	70.83	64.74	35.57	23.61	93.52	35.26	29.17	0.19	65.56
BDI-II 23	70.83	67.31	38.14	25	93.75	32.69	29.17	0.21	67.78
BDI-II 24	66.67	68.59	35.26	24.62	93.04	31.41	33.33	0.2	68.33
BDI-II 25	66.67	69.23	35.9	25	93.1	30.77	33.33	0.21	68.89
BDI-II 27	62.5	75	37.5	27.78	92.86	25	37.5	0.25	73.33
<b>BDI-II 28</b>	<b>58.33</b>	<b>78.21</b>	<b>36.54</b>	<b>29.17</b>	<b>92.42</b>	<b>21.79</b>	<b>41.67</b>	<b>0.26</b>	<b>75.56</b>
BDI-II 29	54.17	80.13	34.3	29.55	91.91	19.87	45.83	0.25	76.67
BDI-II 30	45.83	83.33	29.16	29.73	90.91	16.67	54.17	0.24	78.33
BDI-II 31	37.5	85.26	22.76	28.13	89.86	14.74	62.5	0.2	78.89
BDI-II 32	29.17	88.46	17.63	28	89.03	11.54	70.83	0.17	80.56
BDI-II 33	20.83	90.38	11.21	25	88.13	9.62	79.17	0.12	81.11
BDI-II 34	16.67	91.03	7.7	22.22	87.65	8.97	83.33	0.09	81.11
BDI-II 35	16.67	91.67	8.34	23.53	87.73	8.33	83.33	0.1	81.67

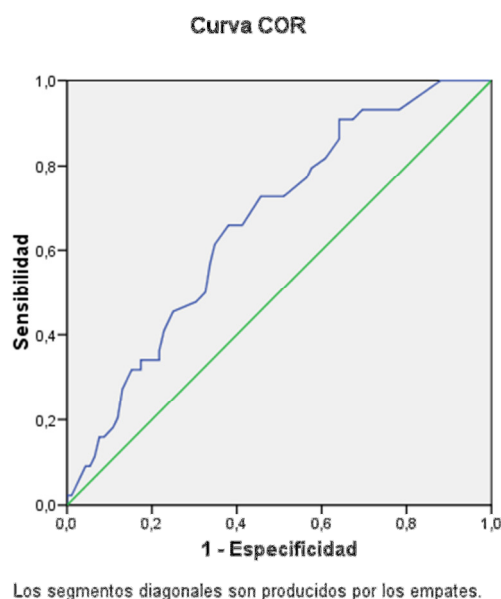


### **Eficacia diagnóstica del MCMI-II y del BDI-II para discriminar entre pacientes con cualquier diagnóstico de trastorno depresivo y pacientes con otro tipo de diagnóstico según el DSM-IV**

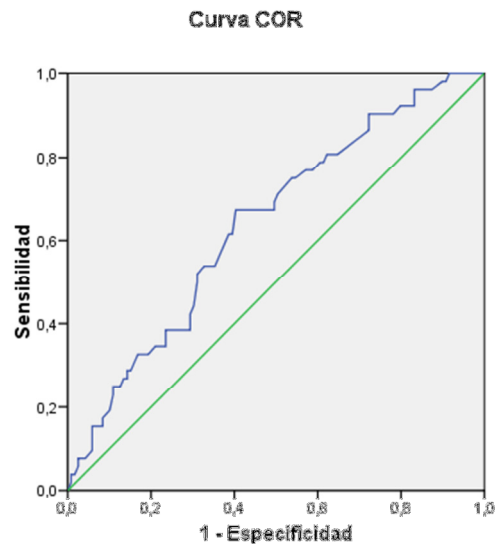
Las curvas ROC realizadas sobre las puntuaciones directas y de tasa base de las escalas CC y D del MCMI-II y sobre las puntuaciones directas del BDI-II, en relación con el diagnóstico de cualquier trastorno depresivo (depresión mayor, distimia, trastorno adaptativo con estado de ánimo deprimido o mixto, duelo y trastorno depresivo no especificado) realizado por los clínicos, se recogen en las Figuras 6.9 a 6.13.

Un análisis visual de las Figuras 6.9 a 6.13 sugería que, en relación con el diagnóstico del conjunto de trastornos depresivos, las curvas ROC de las escalas CC y D del MCMI-II, tanto para sus puntuaciones directas como de TB, y la curva ROC del BDI-II se separaban muy poco de la diagonal que reflejaba un diagnóstico basado en el azar.

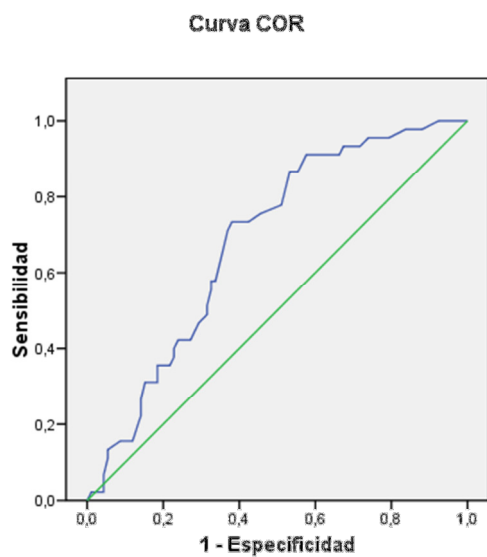
**Figura 6.9.** Curva ROC de las puntuaciones directas de la escala de Depresión Mayor (CC) del MCMI-II con respecto al diagnóstico de cualquier trastorno depresivo



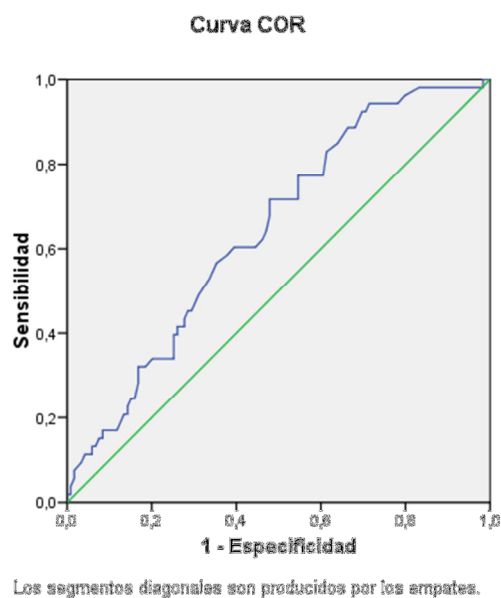
**Figura 6.10.** Curva ROC de las puntuaciones tasa base de la escala de Depresión Mayor (CC) del MCMI-II con respecto al diagnóstico de cualquier trastorno depresivo



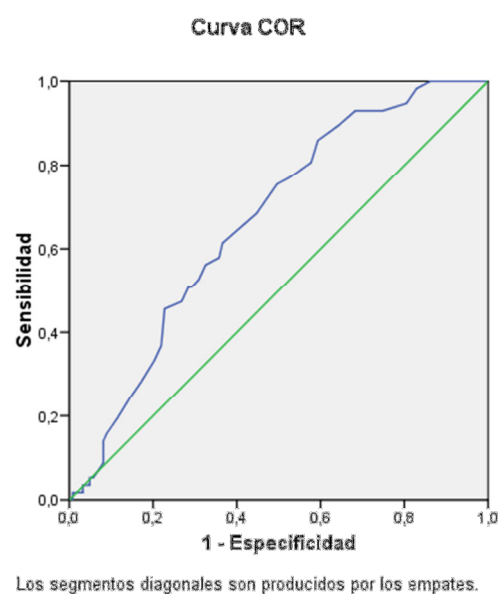
**Figura 6.11.** Curva ROC de las puntuaciones directas de la escala de Distimia (D) del MCMI-II con respecto al diagnóstico de cualquier trastorno depresivo



**Figura 6.12.** Curva ROC de las puntuaciones tasa base de la escala de Distimia (D) del MCMI-II con respecto al diagnóstico de cualquier trastorno depresivo



**Figura 6.13.** Curva ROC de las puntuaciones directas del BDI-II con respecto al diagnóstico de cualquier trastorno depresivo



De hecho, como se puede observar en la Tabla 6.16, el análisis estadístico de las áreas bajo las curvas ROC indicaba, por un lado, que ninguna de esas áreas era estadísticamente diferente, con  $p < 0,05$ , del área de 0.50 que representa un diagnóstico basado en el azar, y, por otro lado, que ninguna de esas áreas alcanzó los criterios mínimos recomendados para un instrumento diagnóstico ( $> 0.70$ , según Swets, 1988, o  $> 0.75$ , según Fan et al., 2006).

**Tabla 6.16.** Áreas bajo la curva ROC en relación con el diagnóstico del conjunto de los trastornos depresivos para las puntuaciones directas y tasa base de las escalas de Depresión Mayor (CC) y Distimia (D) del MCMI-II y de las puntuaciones directas del BDI-II

Puntuación — Escala	Área bajo ROC para diagnóstico trastornos depresivos	Intervalo de confianza asintótico del área al 95%	
		Límite inferior	Límite superior
PD — CC	0.665	0.572	0.759
TB — CC	0.666	0.573	0.760
PD — D	0.679	0.589	0.770
TB — D	0.670	0.579	0.761
BDI-II	0.689	0.599	0.778

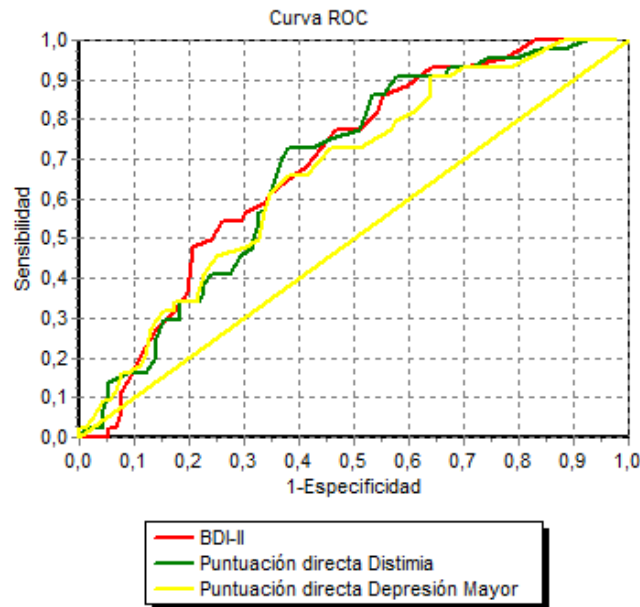
*Nota.* PD — CC = Puntuación directa de la escala de Depresión Mayor del MCMI-II; TB — CC = Puntuación de tasa base de la escala de Depresión Mayor del MCMI-II; PD — D = Puntuación directa de la escala de Distimia del MCMI-II; TB — D = Puntuación de tasa base de la escala de Distimia del MCMI-II; BDI-II = Puntuación directa del Inventario de Depresión de Beck-II.

Igualmente, tanto el análisis visual de las curvas ROC de las Figuras 6.9 a 6.13 como los valores de las áreas de las curvas ROC y de sus intervalos de confianza que se recogen en la Tabla 6.16, sugerían que no existían diferencias significativas en la eficacia diagnóstica de los tres instrumentos, las escalas CC y D del MCMI-II y el BDI-II, para discriminar entre el diagnóstico de cualquier trastorno depresivo y cualquier otro diagnóstico según el DSM-IV.

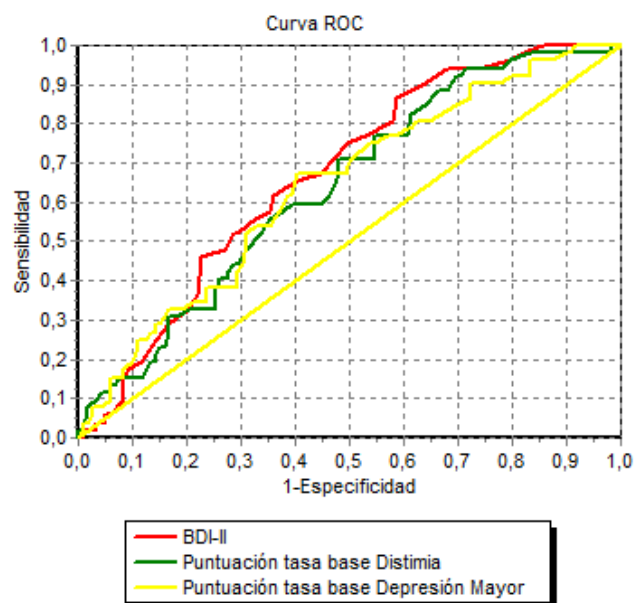
Efectivamente, como puede apreciarse en la Figura 6.14 en la cual se comparan, para el diagnóstico de cualquier trastorno depresivo, las curvas ROC de las puntuaciones directas de las escalas CC y D del MCMI-II y la curva ROC de las puntuaciones directas del BDI-II, estas tres curvas ROC casi se solapaban entre sí. De hecho, el área bajo la curva ROC de las puntuaciones directas de la escala CC no difería significativamente de las áreas bajo la curva ROC de las puntuaciones directas de la escala D ( $z = 0.81$ ,  $p = 0.413$ ) o del BDI-II ( $z = 0.27$ ,  $p = 0.783$ ), ni tampoco difirieran entre sí de forma significativa las áreas bajo la curva ROC de las puntuaciones directas de la escala D y del BDI-II ( $z = 0.64$ ,  $p = 0.516$ ) (véase también el Anexo 9).

Igualmente, como puede observarse en la Figura 6.15 en la cual se comparan, para el diagnóstico de cualquier trastorno depresivo, las curvas ROC de las puntuaciones de TB de las escalas CC y D del MCMI-II y la curva ROC de las puntuaciones directas del BDI-II, estas tres curvas ROC también casi se solapaban entre sí. Es más, el área bajo la curva ROC de las puntuaciones de TB de la escala CC no difería significativamente del área bajo la curva ROC de las puntuaciones de TB de la escala D ( $z = 0.06$ ,  $p = 0.950$ ) o del área bajo la curva ROC de las puntuaciones directas del BDI-II ( $z = 0.84$ ,  $p = 0.396$ ), ni tampoco difirieran entre sí de forma significativa las áreas bajo la curva ROC de las puntuaciones de TB de la escala D y de las puntuaciones directas del BDI-II ( $z = 0.82$ ,  $p = 0.407$ ) (véase también el Anexo 10).

**Figura 6.14.** Curvas ROC de las puntuaciones directas de las escalas de Distimia (D) y de Depresión Mayor (CC) del MCMI-II y de las puntuaciones directas del BDI-II con respecto al diagnóstico del conjunto de trastornos depresivos



**Figura 6.15.** Curvas ROC de las puntuaciones tasa base (TB) de las escalas de Distimia (D) y de Depresión Mayor (CC) del MCMI-II y de las puntuaciones directas del BDI-II con respecto al diagnóstico del conjunto de trastornos depresivos



**Puntos de corte del MCMI-II y del BDI-II para discriminar entre pacientes con cualquier diagnóstico de trastorno depresivo y pacientes con otro tipo de diagnóstico según el DSM-IV**

En las Tablas 6.17, 6.18 y 6.19 se recogen las coordenadas de las curvas ROC de las escalas de CC y D del MCMI-II y del BDI-II respecto al diagnóstico del conjunto de trastornos depresivos, es decir, la sensibilidad y especificidad de cada una de las puntuaciones directas y puntuaciones tasa base (TB) de dichas escalas e instrumentos. En la Tabla 6.17 se puede observar que, con respecto a las puntuaciones directas en la escala CC del MCMI-II, sensibilidades superiores a 0.70 para detectar cualquier trastorno depresivo se alcanzaban con puntuaciones directas inferiores a 16, mientras que especificidades superiores a 0.70 suponían puntuaciones directas superiores a 22. Por otro lado, la Tabla 6.17 también indicaba que sensibilidades superiores a 0.70 para detectar cualquier trastorno depresivo se alcanzaban con puntuaciones TB en la escala de CC del MCMI-II inferiores a 43, mientras que especificidades superiores a 0.70 suponían puntuaciones TB superiores a 59

**Tabla 6.17.** Coordenadas de las curvas ROC de las puntuaciones directas y de tasa base (TB) de la escala de Depresión Mayor (CC) del MCMI-II con respecto al diagnóstico de cualquier trastorno depresivo

Positivo si la puntuación directa de CC es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad	Positivo si la puntuación tasa base de CC es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad
-1.00	1.000	1.000	-1.00	1.000	1.000
0.50	1.000	0.978	1.50	1.000	0.975
1.50	1.000	0.913	3.50	1.000	0.966
2.50	1.000	0.880	4.50	1.000	0.958
3.50	0.955	0.815	5.50	1.000	0.924
4.50	0.932	0.783	7.00	1.000	0.916
5.50	0.932	0.728	8.50	0.981	0.908
6.50	0.932	0.696	9.50	0.981	0.899
7.50	0.909	0.674	11.50	0.962	0.874

Positivo si la puntuación directa de CC es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad	Positivo si la puntuación tasa base de CC es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad
8.50	0.909	0.641	13.50	0.962	0.832
9.50	0.864	0.641	14.50	0.942	0.832
10.50	0.818	0.609	15.50	0.923	0.832
11.50	0.795	0.576	16.50	0.923	0.815
13.00	0.773	0.565	17.50	0.923	0.807
14.50	0.727	0.511	18.50	0.923	0.798
15.50	0.727	0.457	20.50	0.904	0.782
16.50	0.659	0.413	23.00	0.904	0.723
17.50	0.659	0.380	25.00	0.865	0.723
18.50	0.614	0.348	28.00	0.846	0.697
19.50	0.568	0.337	30.50	0.808	0.647
20.50	0.500	0.326	32.50	0.808	0.630
21.50	0.477	0.304	35.00	0.808	0.622
22.50	0.455	0.250	36.50	0.788	0.613
23.50	0.409	0.228	37.50	0.788	0.605
24.50	0.364	0.217	38.50	0.769	0.588
25.50	0.341	0.217	39.50	0.769	0.571
26.50	0.341	0.196	40.50	0.750	0.546
27.50	0.341	0.185	41.50	0.750	0.538
28.50	0.341	0.174	42.50	0.712	0.504
29.50	0.318	0.174	43.50	0.692	0.496
30.50	0.318	0.152	44.50	0.673	0.496
31.50	0.295	0.141	45.50	0.673	0.445
32.50	0.273	0.130	46.50	0.673	0.437
33.50	0.205	0.120	47.50	0.673	0.403
34.50	0.182	0.109	48.50	0.615	0.395
35.50	0.159	0.087	50.00	0.615	0.387
36.50	0.159	0.076	51.50	0.577	0.370
37.50	0.114	0.065	52.50	0.538	0.353
38.50	0.091	0.054	53.50	0.538	0.345
39.50	0.091	0.043	54.50	0.538	0.336
40.50	0.068	0.033	55.50	0.538	0.328
43.00	0.045	0.022	56.50	0.519	0.311
46.50	0.023	0.011	55.50	0.500	0.311
49.50	0.023	0.000	58.50	0.442	0.303
52.00	0.000	0.000	59.50	0.423	0.294
			60.50	0.385	0.294
			61.50	0.385	0.286
			62.50	0.385	0.277
			63.50	0.385	0.252
			64.50	0.385	0.244
			65.50	0.385	0.235
			67,00	0.346	0.235



Positivo si la puntuación directa de CC es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad	Positivo si la puntuación tasa base de CC es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad
			68.50	0.346	0.227
			69.50	0.346	0.218
			70.50	0.346	0.210
			71.50	0.327	0.193
			72.50	0.327	0.168
			74.00	0.308	0.160
			75.50	0.288	0.151
			77.50	0.288	0.143
			79.50	0.269	0.143
			80.50	0.269	0.134
			81.50	0.250	0.126
			83.00	0.250	0.118
			84.50	0.250	0.109
			86.00	0.231	0.109
			88.00	0.192	0.101
			89.50	0.173	0.084
			90.50	0.154	0.084
			91.50	0.154	0.076
			93.00	0.154	0.059
			94.50	0.115	0.059
			95.50	0.096	0.059
			96.50	0.077	0.042
			100.00	0.077	0.025
			104.00	0.058	0.025
			105.50	0.038	0.017
			107.50	0.038	0.008
			109.50	0.019	0.008
			111.00	0.000	0.000

*Nota.* <sup>a</sup>El punto de corte menor es la mínima puntuación observada menos 1, mientras que el punto de corte mayor es la máxima puntuación observada más 1. Todos los demás puntos de corte son la media de dos puntuaciones observadas ordenadas y consecutivas. CC = Escala de Depresión Mayor del MCMI-II.

En la Tabla 6.18 se recogen las coordenadas de las curvas ROC de las puntuaciones directas y de tasa base (TB) de la escala de D del MCMI-II con respecto al diagnóstico del conjunto de trastornos depresivos, las cuales indicaban que sensibilidades superiores a 0.70 para detectar cualquier trastorno depresivo se alcanzaban con puntuaciones directas y TB inferiores a 25 y 52, respectivamente,

mientras que especificidades superiores a 0.70 suponían puntuaciones directas y TB superiores a 30 y 70, respectivamente.

**Tabla 6.18.** Coordenadas de las curvas ROC de las puntuaciones directas y de tasa base (TB) de la escala de Distimia (D) del MCMI-II con respecto al diagnóstico de cualquier trastorno depresivo

Positivo si la puntuación directa de D es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad	Positivo si la puntuación tasa base de D es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad
-1.00	1.000	1.000	-1.00	1.000	1.000
0.50	1.000	0.989	4.00	1.000	0.992
1.50	1.000	0.978	10.00	1.000	0.983
2.50	1.000	0.946	12.50	0.981	0.983
3.50	1.000	0.924	14.50	0.981	0.975
4.50	0.978	0.880	16.50	0.981	0.966
5.50	0.978	0.837	18.00	0.981	0.958
6.50	0.956	0.793	19.50	0.981	0.941
7.50	0.956	0.739	21.00	0.981	0.916
8.50	0.933	0.717	22.50	0.981	0.899
9.50	0.933	0.696	23.50	0.981	0.882
10.50	0.933	0.674	24.50	0.981	0.857
11.50	0.911	0.663	26.00	0.981	0.849
12.50	0.911	0.641	27.50	0.981	0.840
13.50	0.911	0.598	28.50	0.981	0.832
14.50	0.911	0.576	29.50	0.962	0.798
15.50	0.867	0.554	30.50	0.943	0.782
16.50	0.867	0.543	32.00	0.943	0.773
17.50	0.867	0.533	33.50	0.943	0.731
18.50	0.822	0.522	34.50	0.943	0.714
19.50	0.778	0.511	35.50	0.925	0.706
20.50	0.756	0.457	37.00	0.925	0.697
21.50	0.733	0.424	38.50	0.887	0.681
22.50	0.733	0.391	39.50	0.887	0.672
23.50	0.733	0.380	40.50	0.887	0.664
24.50	0.711	0.370	41.50	0.849	0.639
25.50	0.578	0.337	42.50	0.830	0.613
26.50	0.578	0.326	43.50	0.774	0.605
27.50	0.556	0.326	44.50	0.774	0.571
28.50	0.511	0.315	45.50	0.774	0.563
29.50	0.489	0.315	46.50	0.774	0.546
30.50	0.467	0.293	47.50	0.717	0.546
31.50	0.422	0.272	48.50	0.717	0.538
32.50	0.422	0.250	49.50	0.717	0.521

Positivo si la puntuación directa de D es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad	Positivo si la puntuación tasa base de D es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad
33.50	0.422	0.239	50.50	0.717	0.496
34.50	0.400	0.228	51.50	0.717	0.479
35.50	0.378	0.228	52.50	0.679	0.479
37.00	0.356	0.217	54.00	0.642	0.471
39.00	0.356	0.196	56.00	0.623	0.462
40.50	0.356	0.185	57.50	0.604	0.445
41.50	0.333	0.185	58.50	0.604	0.420
42.50	0.311	0.185	59.50	0.604	0.412
43.50	0.311	0.152	61.00	0.604	0.403
44.50	0.267	0.141	62.50	0.604	0.395
45.50	0.222	0.141	63.50	0.585	0.378
47.00	0.156	0.120	64.50	0.566	0.353
49.00	0.156	0.087	66.00	0.528	0.336
50.50	0.133	0.054	68.00	0.491	0.311
51.50	0.111	0.054	69.50	0.472	0.303
53.00	0.067	0.043	70.50	0.453	0.294
54.50	0.022	0.043	71.50	0.453	0.286
55.50	0.022	0.022	72.50	0.434	0.277
58.00	0.022	0.011	73.50	0.415	0.277
61.00	0.000	0.000	74.50	0.415	0.269
			75.50	0.415	0.261
			76.50	0.396	0.261
			77.50	0.396	0.252
			78.50	0.340	0.252
			80.00	0.340	0.244
			81.50	0.340	0.235
			82.50	0.340	0.227
			83.50	0.340	0.202
			84.50	0.321	0.185
			85.50	0.321	0.168
			86.50	0.302	0.168
			87.50	0.283	0.168
			88.50	0.245	0.160
			90.00	0.245	0.151
			91.50	0.226	0.143
			92.50	0.208	0.143
			93.50	0.208	0.134
			94.50	0.170	0.118
			95.50	0.170	0.101
			96.50	0.170	0.084
			97.50	0.151	0.084
			98.50	0.151	0.076

Positivo si la puntuación directa de D es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad	Positivo si la puntuación tasa base de D es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad
			100.50	0.132	0.067
			103.00	0.132	0.059
			104.50	0.113	0.059
			105.50	0.113	0.042
			106.50	0.094	0.034
			107.50	0.075	0.017
			108.50	0.057	0.017
			109.50	0.038	0.008
			111.50	0.019	0.008
			118.50	0.019	0.000
			125.00	0.000	0.000

*Nota.* <sup>a</sup>El punto de corte menor es la mínima puntuación observada menos 1, mientras que el punto de corte mayor es la máxima puntuación observada más 1. Todos los demás puntos de corte son la media de dos puntuaciones observadas ordenadas y consecutivas. D MCMI-II = Escala de Distimia del MCMI-II.

Finalmente, en la Tabla 6.19 se recogen las coordenadas de la curva ROC de las puntuaciones directas del BDI-II con respecto al diagnóstico del conjunto de trastornos depresivos, las cuales indicaban que sensibilidades superiores a 0.70 para detectar cualquier trastorno depresivo se alcanzaban con puntuaciones en el BDI-II inferiores a 17, mientras que especificidades superiores a 0.70 suponían puntuaciones en el BDI-II superiores a 23.

**Tabla 6.19.** Coordenadas de las curvas ROC de las puntuaciones directas del BDI-II con respecto al diagnóstico de cualquier trastorno depresivo

Positivo si la puntuación del BDI-II es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad
-1.00	1.000	1.000
0.50	1.000	0.984
1.50	1.000	0.976
2.50	1.000	0.967
3.50	1.000	0.951
4.50	1.000	0.927
5.50	1.000	0.886
6.50	1.000	0.862
7.50	0.982	0.829

Positivo si la puntuación del BDI-II es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad
8.50	0.947	0.805
9.50	0.930	0.748
10.50	0.930	0.732
11.50	0.930	0.683
12.50	0.895	0.642
13.50	0.860	0.593
14.50	0.807	0.577
15.50	0.772	0.528
16.50	0.754	0.496
17.50	0.684	0.447
18.50	0.649	0.407
19.50	0.614	0.366
20.50	0.579	0.358
21.50	0.561	0.325
22.50	0.526	0.309
23.50	0.509	0.293
24.50	0.509	0.285
25.50	0.474	0.268
26.50	0.456	0.228
27.50	0.368	0.220
28.50	0.333	0.203
29.50	0.281	0.171
30.50	0.246	0.146
31.50	0.193	0.114
32.50	0.158	0.089
33.50	0.140	0.081
34.50	0.123	0.081
35.50	0.105	0.081
36.50	0.088	0.081
37.50	0.053	0.057
38.50	0.053	0.049
39.50	0.035	0.049
40.50	0.035	0.041
41.50	0.035	0.033
42.50	0.018	0.033
44.00	0.018	0.008
47.50	0.000	0.008
51.00	0.000	0.000

*Nota.* <sup>a</sup>El punto de corte menor es la mínima puntuación observada menos 1, mientras que el punto de corte mayor es la máxima puntuación observada más 1. Todos los demás puntos de corte son la media de dos puntuaciones observadas ordenadas y consecutivas.

En las Tablas 6.20, 6.21 y 6.22, además de la sensibilidad y la especificidad, se recoge un mayor número de índices de validez diagnóstica (índice de Youden, poder predictivo positivo, poder predictivo negativo, eficacia global y kappa) en relación con el diagnóstico del conjunto de trastornos depresivos para aquellas puntuaciones de las escalas CC y D del MCMI-III y del BDI-II que han sido propuestas en los manuales correspondientes como puntos de corte para el diagnóstico de los trastornos depresivos (TB 75 y TB 85 del MCMI-III) o para la delimitación de diferentes niveles de gravedad de la sintomatología depresiva (BDI-II = 14, 20 y 28). Así mismo, se presentan esos mismos índices de validez diagnóstica para algunas puntuaciones que se encuentran cercanas a las anteriores.

Como se puede comprobar en la Tabla 6.20, los puntos de corte propuestos por los autores del MCMI-III para la escala CC (TB 75 y TB 85) no son los que obtuvieron los valores más elevados de concordancia con el diagnóstico de referencia (kappa) y de equilibrio entre sensibilidad y especificidad (índice de Youden) en el caso del diagnóstico del conjunto de trastornos depresivos. Por ejemplo, la puntuación TB 75 tuvo un índice de Youden bajo de 14.8, con una alta especificidad (84.03%), pero una pobre sensibilidad (30.77%) y un índice kappa pobre de 0.16. Respecto a las puntuaciones directas de la escala CC, la puntuación 9 tuvo el mayor índice de Youden (26.78), con una sensibilidad muy alta (90.91%), una muy baja especificidad (35.87%) y un índice kappa superior (0.2), pero todavía en el rango de valores de kappa considerados pobres. Por el contrario, el punto de corte 18 tuvo un mayor equilibrio entre sensibilidad y especificidad (65.91% y 61.96%, respectivamente), con un índice de Youden ligeramente inferior (27.87) y un índice kappa ligeramente superior de 0.25, que puede considerarse débil (Altman, 1991).

Los índices de validez para el conjunto de trastornos depresivos de los puntos de corte habituales de la escala D del MCMI-II (TB 75 y TB 85) no fueron tampoco los que

obtuvieron los valores más elevados de concordancia con el diagnóstico de referencia (kappa) y de equilibrio entre sensibilidad y especificidad (índice de Youden) (véase la Tabla 6.21). En concreto, los puntos de corte TB 61 y TB 62 fueron los que tuvieron los mayores índices kappa (0.18) y de Youden (20.04), con una sensibilidad de 60.38% y una especificidad de 59.66%, mientras que las puntuaciones TB 75 y TB 85 tuvieron, respectivamente, índices kappa de 0.15, sensibilidades de 41.51% y 32.08%, especificidades de 73.11% y 81.51%, e índices de Youden de 14.62 y 13.59. No obstante, en todos los casos (TB 61, TB 62, TB 75 y TB 85), los índices kappa fueron pobres y los índices de sensibilidad, especificidad y de Youden estuvieron por debajo de los valores aceptables (salvo por los índices de especificidad de las puntuaciones TB 75 y TB 85; véase la Tabla 6.21).

La puntuación directa de la escala D del MCMI-II que tuvo un índice de Youden de más elevado a la hora de discriminar el conjunto de trastornos depresivos frente a otros trastornos mentales fue la puntuación de 23 (34.2), con una alta sensibilidad (73.33%) a costa de una menor especificidad (60.87%), y un el mayor índice kappa, pero también débil (0.3). La puntuación directa de 15 tuvo un índice kappa ligeramente inferior (26), con un índice de Youden menor (33.5) y con una alta sensibilidad (91.11%) y una menor especificidad (42.39%) (véase la Tabla 6.21). De forma similar a lo que ocurría con las puntuaciones TB, los índices kappa de esas dos puntuaciones directas fueron débiles y sus índices de especificidad y de Youden estuvieron, en general, por debajo de los valores aceptables, salvo el índice de sensibilidad de las puntuaciones directas (> 70%).

En cuanto la eficacia para el diagnóstico del conjunto de trastornos depresivos de las puntuaciones directas del BDI-II (véase Tabla 6.22), la puntuación 14 obtuvo el mayor índice de Youden (26.61). Este punto de corte tuvo una alta sensibilidad (85.96%), pero una especificidad no aceptable (40.65%) y un índice kappa pobre (0.2). De hecho, su

poder predictivo positivo fue insuficiente (40.16%), aunque su poder predictivo negativo fue muy alto (86.21%). Con un índice de Youden menor (22.85), la puntuación de 27 en el BDI-II mostró una especificidad aceptable (77.24%) y un índice kappa mayor (0.23), pero aun así pobre; es más, su sensibilidad tampoco fue aceptable (45.61%).

Observando las tres tablas que presentan los índices de eficacia diagnóstica de diferentes puntuaciones de las escalas CC y D del MCMI-II y del BDI-II para el diagnóstico del conjunto de trastornos depresivos (Tablas 6.20, 6.21 y 6.22), se puede constatar que, al igual que se comentó con respecto al diagnóstico del trastorno depresivo mayor, el poder predictivo positivo de los instrumentos para diagnosticar un trastorno depresivo de cualquier tipo es menor que su poder predictivo negativo, aunque los índices de poder predictivo positivo parecen, en general, ligeramente superiores en el caso del diagnóstico del conjunto de trastornos depresivos que en el caso del diagnóstico del trastorno depresivo mayor.



**Tabla 6.20.** Índices de validez diagnóstica de distintas puntuaciones directas (PD) o tasa base (TB) de la escala de Depresión Mayor (CC) del MCMI-II con respecto al diagnóstico del conjunto de trastornos depresivos

Punto de corte – Escala	Sensibilidad	Especificidad	Índice de Youden	Poder predictivo positivo	Poder predictivo negativo	Tasa de falsos positivos	Tasa de falsos negativos	Kappa	Eficacia
<b>TB 75</b>	<b>30.77</b>	<b>84.03</b>	<b>14.8</b>	<b>45.71</b>	<b>73.53</b>	<b>15.97</b>	<b>69.23</b>	<b>0.16</b>	<b>67.84</b>
<b>TB 85</b>	<b>25</b>	<b>89.08</b>	<b>14.08</b>	<b>50</b>	<b>73.1</b>	<b>10.92</b>	<b>75</b>	<b>0.16</b>	<b>69.59</b>
TB 66	38.46	76.47	14.93	41.67	73.98	23.53	61.54	0.15	64.91
TB 67	34.62	76.47	11.09	39.13	72.8	23.53	65.38	0.11	63.74
TB 68	34.62	76.47	11.09	39.13	72.8	23.53	65.38	0.11	63.74
TB 62	38.46	71.43	9.89	37.04	72.65	28.57	61.54	0.1	61.4
TB 61	38.46	70.59	9.05	36.36	72.41	29.41	61.54	0.09	60.82
TB 60	42.31	70.59	12.9	38.6	73.68	29.41	57.69	0.13	61.99
PD 7	93.18	30.43	23.61	39.05	90.32	69.57	6.82	0.17	50.74
<b>PD 9</b>	<b>90.91</b>	<b>35.87</b>	<b>26.78</b>	<b>40.4</b>	<b>89.19</b>	<b>64.13</b>	<b>9.09</b>	<b>0.2</b>	<b>53.68</b>
PD 8	90.91	32.61	23.52	39.22	88.24	67.39	9.09	0.18	51.47
PD 10	86.36	35.87	22.23	39.18	84.62	64.13	13.64	0.17	52.21
PD 11	81.82	39.13	20.95	39.13	81.82	60.87	18.18	0.16	52.94
PD 12	79.55	42.39	21.94	39.77	81.25	57.61	20.45	0.17	54.41
PD 13	77.27	43.48	20.75	39.53	80	56.52	22.73	0.17	54.41
PD 14	77.27	43.48	20.75	39.53	80	56.52	22.73	0.17	54.41
PD 15	72.73	48.91	21.64	40.51	78.95	51.09	27.27	0.18	56.62
PD 16	72.73	54.35	27.08	43.24	80.65	45.65	27.27	0.23	60.29
PD 17	65.91	58.7	24.61	43.28	78.26	41.3	34.09	0.22	61.03
<b>PD 18</b>	<b>65.91</b>	<b>61.96</b>	<b>27.87</b>	<b>45.31</b>	<b>79.17</b>	<b>38.04</b>	<b>34.09</b>	<b>0.25</b>	<b>63.24</b>
PD 19	61.36	65.22	26.58	45.76	77.92	34.78	38.64	0.24	63.97
PD 20	56.82	66.3	23.12	44.64	76.25	33.7	43.18	0.22	63.24
PD 23	45.45	75	20.45	46.51	74.19	25	54.55	0.21	65.44
PD 28	34.09	81.52	15.61	46.88	72.12	18.48	65.91	0.17	66.18
PD 32	29.55	85.87	15.42	50	71.82	14.13	70.45	0.17	67.65
PD 35	18.18	89.13	7.31	44.44	69.49	10.87	81.82	0.09	66.18

**Tabla 6.21.** Índices de validez diagnóstica de distintas puntuaciones directas (PD) o tasa base (TB) de la escala de Distimia (D) del MCMI-II con respecto al diagnóstico del conjunto de trastornos depresivos

Punto de corte – Escala	Sensibilidad	Especificidad	Índice de Youden	Poder predictivo positivo	Poder predictivo negativo	Tasa de falsos positivos	Tasa de falsos negativos	Kappa	Eficiencia
<b>TB 85</b>	<b>32.08</b>	<b>81.51</b>	<b>13.59</b>	<b>43.59</b>	<b>72.93</b>	<b>18.49</b>	<b>67.92</b>	<b>0.15</b>	<b>66.28</b>
<b>TB 75</b>	<b>41.51</b>	<b>73.11</b>	<b>14.62</b>	<b>40.74</b>	<b>73.73</b>	<b>26.89</b>	<b>58.49</b>	<b>0.15</b>	<b>63.37</b>
TB 66	52.83	66.39	19.22	41.18	75.96	33.61	47.17	0.18	62.21
TB 67	52.83	66.39	19.22	41.18	75.96	33.61	47.17	0.18	62.21
TB 68	49.06	68.91	17.97	41.27	75.23	31.09	50.94	0.17	62.79
TB 60	60.38	58.82	19.2	39.51	76.92	41.18	39.62	0.17	59.3
<b>TB 61</b>	<b>60.38</b>	<b>59.66</b>	<b>20.04</b>	<b>40</b>	<b>77.17</b>	<b>40.34</b>	<b>39.62</b>	<b>0.18</b>	<b>59.88</b>
<b>TB 62</b>	<b>60.38</b>	<b>59.66</b>	<b>20.04</b>	<b>40</b>	<b>77.17</b>	<b>40.34</b>	<b>39.62</b>	<b>0.18</b>	<b>59.88</b>
PD 8	95.56	26.09	21.65	38.74	92.31	73.91	4.44	0.16	48.91
PD 9	93.33	28.26	21.59	38.89	89.66	71.74	6.67	0.16	49.64
PD 10	93.33	30.43	23.76	39.62	90.32	69.57	6.67	0.18	51.09
PD 11	93.33	32.61	25.94	40.38	90.91	67.39	6.67	0.19	52.55
<b>PD 15</b>	<b>91.11</b>	<b>42.39</b>	<b>33.5</b>	<b>43.62</b>	<b>90.7</b>	<b>57.61</b>	<b>8.89</b>	<b>0.26</b>	<b>58.39</b>
PD 18	86.67	46.74	33.41	44.32	87.76	53.26	13.33	0.27	59.85
PD 20	77.78	48.91	26.69	42.68	81.82	51.09	22.22	0.22	58.39
<b>PD 23</b>	<b>73.33</b>	<b>60.87</b>	<b>34.2</b>	<b>47.83</b>	<b>82.35</b>	<b>39.13</b>	<b>26.67</b>	<b>0.3</b>	<b>64.96</b>
PD 28	55.56	67.39	22.95	45.45	75.61	32.61	44.44	0.22	63.5
PD 32	42.22	72.83	15.05	43.18	72.04	27.17	57.78	0.15	62.77
PD 35	40	77.17	17.17	46.15	72.45	22.83	60	0.18	64.96

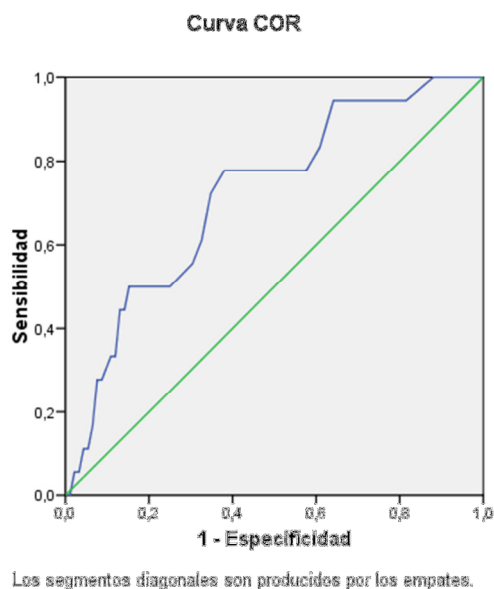
**Tabla 6.22.** Índices de validez diagnóstica de distintas puntuaciones del BDI-II con respecto al diagnóstico del conjunto de trastornos depresivos

Punto de corte — Escala	Sensibilidad	Especificidad	Índice de Youden	Poder predictivo positivo	Poder predictivo negativo	Tasa de falsos positivos	Tasa de falsos negativos	Kappa	Eficiencia
<b>BDI-II 14</b>	<b>85.96</b>	<b>40.65</b>	<b>26.61</b>	<b>40.16</b>	<b>86.21</b>	<b>59.35</b>	<b>14.04</b>	<b>0.2</b>	<b>55</b>
BDI-II 15	80.7	42.28	22.98	39.32	82.54	57.72	19.3	0.18	54.44
BDI-II 16	77.19	47.15	24.34	40.37	81.69	52.85	22.81	0.2	56.67
BDI-II 17	75.44	50.41	25.85	41.35	81.58	49.59	24.56	0.21	58.33
BDI-II 18	68.42	55.28	23.7	41.49	79.07	44.72	31.58	0.2	59.44
BDI-II 19	64.91	59.35	24.26	42.53	78.49	40.65	35.09	0.21	61.11
BDI-II 20	61.4	63.41	24.81	43.75	78	36.59	38.6	0.22	62.78
BDI-II 21	57.89	64.23	22.12	42.86	76.7	35.77	42.11	0.2	62.22
BDI-II 22	56.14	67.48	23.62	44.44	76.85	32.52	43.86	0.22	63.89
BDI-II 23	52.63	69.11	21.74	44.12	75.89	30.89	47.37	0.21	63.89
BDI-II 24	50.88	70.73	21.61	44.62	75.65	29.27	49.12	0.21	64.44
BDI-II 25	50.88	71.54	22.42	45.31	75.86	28.46	49.12	0.22	65
<b>BDI-II 27</b>	<b>45.61</b>	<b>77.24</b>	<b>22.85</b>	<b>48.15</b>	<b>75.4</b>	<b>22.76</b>	<b>54.39</b>	<b>0.23</b>	<b>67.22</b>
BDI-II 28	36.84	78.05	14.89	43.75	72.73	21.95	63.16	0.16	65
BDI-II 29	33.33	79.67	13	43.18	72.06	20.33	66.67	0.14	65
BDI-II 30	28.07	82.93	11	43.24	71.33	17.07	71.93	0.12	65.56
BDI-II 31	24.56	85.37	9.93	43.75	70.95	14.63	75.44	0.11	66.11
BDI-II 32	19.3	88.62	7.92	44	70.32	11.38	80.7	0.09	66.67
BDI-II 33	15.79	91.06	6.85	45	70	8.94	84.21	0.08	67.22
BDI-II 34	14.04	91.87	5.91	44.44	69.75	8.13	85.96	0.07	67.22
BDI-II 35	12.28	91.87	4.15	41.18	69.33	8.13	87.72	0.05	66.67

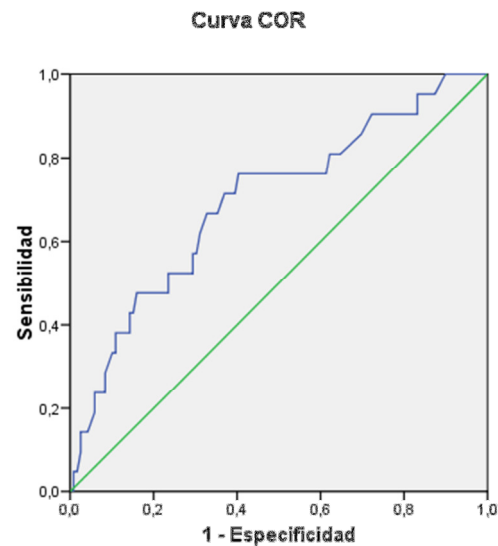
**Eficacia diagnóstica de las escalas CC y D del MCMI-II y del BDI-II para discriminar entre pacientes con un diagnóstico de trastorno depresivo mayor y pacientes con otro tipo de trastorno mental no depresivo según el DSM-IV**

Siguiendo las sugerencias de Ganellen (1996) señaladas al inicio de este capítulo, se realizó también el análisis de las curvas ROC para el diagnóstico del trastorno depresivo mayor frente a otros tipos de trastornos mentales no depresivos, es decir, excluyendo entre estos últimos el resto de trastornos depresivos (distimia, trastorno del estado de ánimo no especificado, duelo y trastorno adaptativo con estado de ánimo depresivo o mixto). Las curvas ROC de las escalas CC y D del MCMI-II y del BDI-II para esta discriminación diagnóstica se presentan en las Figuras 6.16, 6.17, 6.18, 6.19 y 6.20.

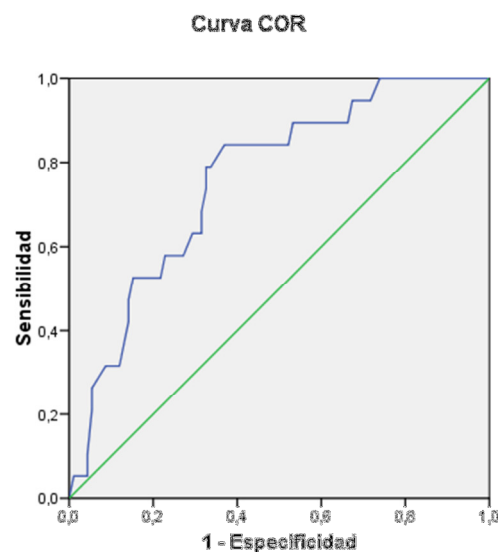
**Figura 6.16.** Curva ROC de las puntuaciones directas de la escala de Depresión Mayor (CC) del MCMI-II con respecto al diagnóstico de trastorno depresivo mayor (pacientes con el diagnóstico de otros trastornos depresivos excluidos)



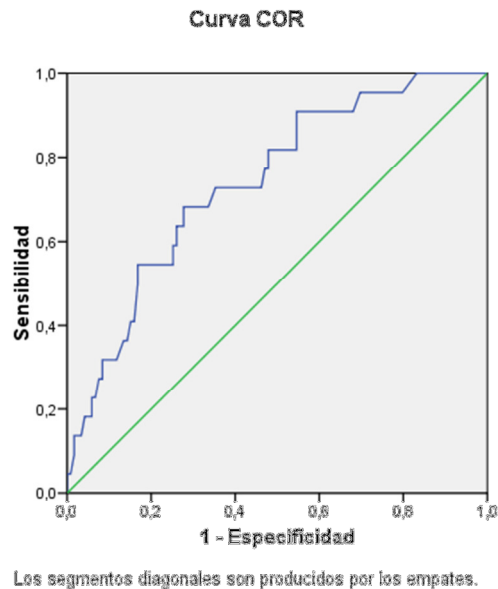
**Figura 6.17.** Curva ROC de las puntuaciones tasa base (TB) de la escala de Depresión Mayor (CC) del MCMI-II con respecto al diagnóstico trastorno depresivo mayor (pacientes con el diagnóstico de otros trastornos depresivos excluidos)



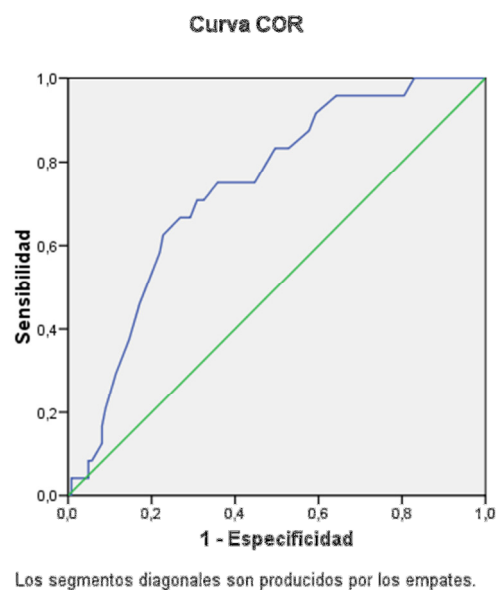
**Figura 6.18.** Curva ROC de las puntuaciones directas de la escala de Distimia (D) del MCMI-II con respecto al diagnóstico de trastorno depresivo mayor (pacientes con el diagnóstico de otros trastornos depresivos excluidos)



**Figura 6.19.** Curva ROC de las puntuaciones tasa base (TB) de la escala de Distimia (D) del MCMI-II con respecto al diagnóstico de trastorno depresivo mayor (pacientes con el diagnóstico de otros trastornos depresivos excluidos)



**Figura 6.20.** Curva ROC de las puntuaciones del BDI-II con respecto al diagnóstico de trastorno depresivo mayor (pacientes con el diagnóstico de otros trastornos depresivos excluidos)



Se puede observar a través de un análisis visual de las Figuras 6.16 a 6.20 que las áreas bajo la curva ROC de las puntuaciones directas o tasas base de las escalas CC y D del MCMI-II y de las puntuaciones del BDI-II se separan moderadamente de los segmentos diagonales producidos por los empates entre sensibilidad y la tasa de falsos positivos, es decir, del diagnóstico basado en el azar (representado en las figuras mediante líneas diagonales verdes). De hecho, como se puede observar en la Tabla 6.23, el análisis estadístico de las áreas bajo las curvas ROC indicaba que todas las áreas eran estadísticamente diferentes, con  $p < 0,05$ , del área de 0.50 que representa un diagnóstico basado en el azar, y todas esas áreas alcanzaron el criterio mínimo recomendado por algunos investigadores para un instrumento diagnóstico ( $> 0.70$ , según Swets, 1988). Además, las puntuaciones directas y tasa base de la escala CC y las puntuaciones del BDI-II alcanzaron el criterio más estricto propuesto por otros investigadores ( $> 0.75$ , según Fan et al., 2006).

Sin embargo, el análisis visual de la Figura 6.20 no sugiere que el área bajo la curva ROC de las puntuaciones del BDI-II y las puntuaciones directas y tasa base de la escala CC del MCMI-II se separen de la línea diagonal verde que representa un diagnóstico basado al azar en mayor medida que lo hacían las áreas bajo las curvas ROC de las puntuaciones directas y tasa base de la escala D del MCMI-II y, de hecho, existían solapamientos entre los intervalos de confianza al 95% de las áreas de las curvas ROC de los tres instrumentos (véase la Tabla 6.23) que sugerían que sus curvas no eran diferentes de forma estadísticamente significativa.

**Tabla 6.23.** Áreas bajo la curva ROC en relación con el diagnóstico de trastorno depresivo mayor (pacientes con el diagnóstico de otros trastornos depresivos excluidos) para las puntuaciones directas y tasa base de las escalas de Depresión Mayor (CC) y Distimia (D) del MCMI-II y para las puntuaciones directas del BDI-II

Puntuación — Escala	Área bajo curva ROC para el diagnóstico de trastorno depresivo mayor	Intervalo de confianza asintótico del área al 95%	
		Límite inferior	Límite superior
PD — CC	0.754*	0.641	0.866
TB — CC	0.752*	0.647	0.858
PD — D	0.718*	0.594	0.843
TB — D	0.727*	0.604	0.851
BDI-II	0.782*	0.685	0.879

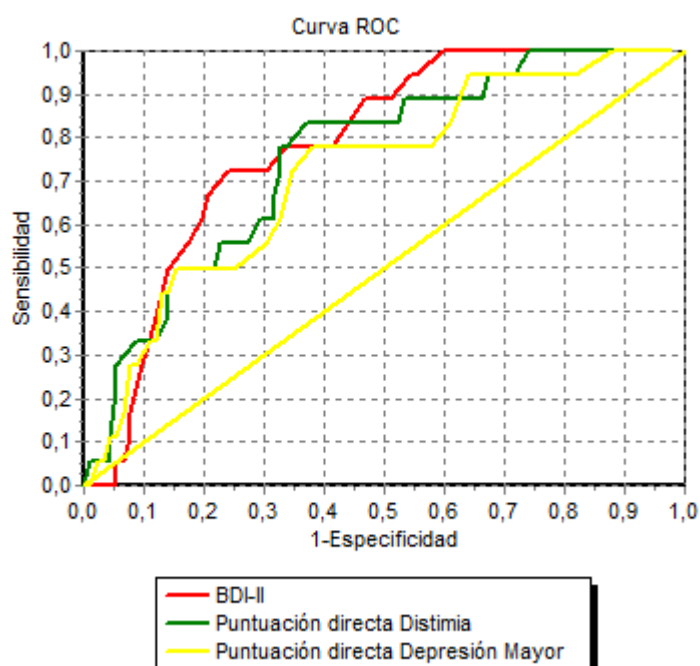
*Nota.* \* Estadísticamente significativa con  $p < 0,001$ . PD — CC = Puntuación directa de la escala de Depresión Mayor del MCMI-II; TB — CC = Tasa base de la escala de Depresión Mayor del MCMI-II; PD — D = Puntuación directa de la escala de Distimia del MCMI-II; TB — D = Tasa base de la escala de Distimia del MCMI-II; BDI-II = Puntuación directa del Inventario de Depresión de Beck-II.

Efectivamente, como puede apreciarse en la Figura 6.21 en la cual se comparan, para el diagnóstico de trastorno depresivo mayor frente al diagnóstico de un trastorno mental no depresivo, las curvas ROC de las puntuaciones directas de las escalas CC y D del MCMI-II y la curva ROC de las puntuaciones directas del BDI-II, estas tres curvas ROC casi se solapaban entre sí. De hecho, el área bajo la curva ROC de las puntuaciones directas de la escala CC no difería significativamente de las áreas bajo la curva ROC de las puntuaciones directas de la escala D ( $z = 1.46$ ,  $p = 0.143$ ) o del BDI-II



( $z = 0.55$ ,  $p = 0.580$ ), ni tampoco difirieran entre sí de forma significativa las áreas bajo la curva ROC de las puntuaciones directas de la escala D y del BDI-II ( $z = 1.24$ ,  $p = 0.214$ ) (véase también el Anexo 11).

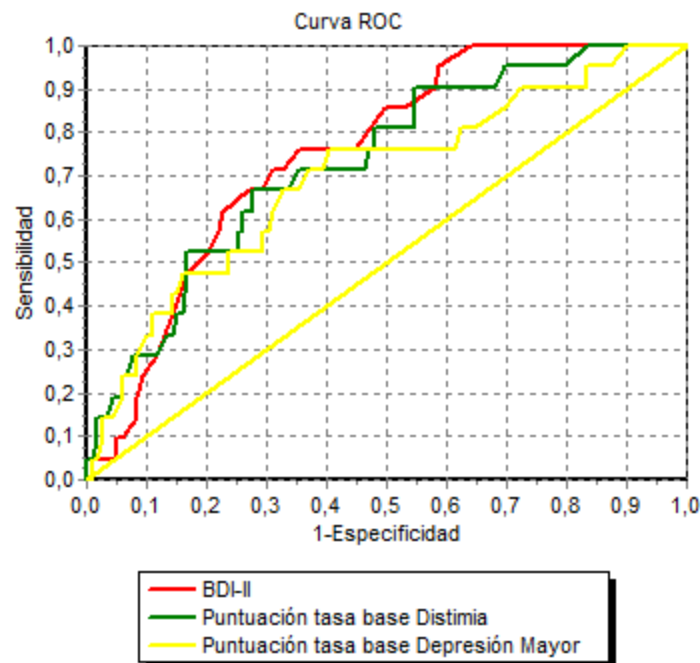
**Figura 6.21.** Curvas ROC de las puntuaciones directas de la escala de Distimia y Depresión Mayor del MCMI-II con respecto al diagnóstico de Trastorno Depresivo Mayor (pacientes con el diagnóstico de otros trastornos depresivos excluidos)



Igualmente, como puede observarse en la Figura 6.22 en la cual se comparan, para el diagnóstico de trastorno depresivo mayor frente al diagnóstico de un trastorno mental no depresivo, las curvas ROC de las puntuaciones de TB de las escalas CC y D del MCMI-II y la curva ROC de las puntuaciones directas del BDI-II, estas tres curvas ROC también casi se solapaban entre sí. Es más, el área bajo la curva ROC de las puntuaciones de TB de la escala CC no difería significativamente del área bajo la curva ROC de las puntuaciones de TB de la escala D ( $z = 1.18$ ,  $p = 0.235$ ) o del área bajo la

curva ROC de las puntuaciones directas del BDI-II ( $z = 0.47$ ,  $p = 0.634$ ), ni tampoco difirieran entre sí de forma significativa las áreas bajo la curva ROC de las puntuaciones de TB de la escala D y de las puntuaciones directas del BDI-II ( $z = 1.13$ ,  $p = 0.258$ ) (véase también el Anexo 12).

**Figura 6.22.** Curvas ROC de las puntuaciones tasa base de las escalas de Distimia y Depresión Mayor del MCMI-II con respecto al diagnóstico de trastorno depresivo mayor (pacientes con el diagnóstico de otros trastornos depresivos excluidos)



**Puntos de corte del MCMI-II y del BDI-II para discriminar entre pacientes con un diagnóstico de trastorno depresivo mayor y pacientes con otro tipo de trastorno mental no depresivo según el DSM-IV**

En las Tablas 6.24, 6.25 y 6.26 se recogen las coordenadas de las curvas ROC de las escalas de CC y D del MCMI-II y del BDI-II respecto al diagnóstico del trastorno depresivo mayor (pacientes con el diagnóstico de otros trastornos depresivos excluidos). Estas coordenadas representan la sensibilidad y especificidad específicas de cada una

de las puntuaciones directas y puntuaciones tasa base (TB) de las escalas. Por ejemplo, en la Tabla 6.24 se puede observar que sensibilidades superiores a 0.70 para el diagnóstico del trastorno depresivo mayor se alcanzaban con puntuaciones directas y TB en la escala de CC del MCMI-II inferiores a 19 y 52, respectivamente, mientras que especificidades superiores a 0.70 suponían puntuaciones directas y TB superiores a 22 y 59, respectivamente.

**Tabla 6.24.** Coordenadas de las curvas ROC de las puntuaciones directas y de tasa base (TB) de la escala de Depresión Mayor (CC) del MCMI-II con respecto al diagnóstico de trastorno depresivo mayor (pacientes con el diagnóstico de otros trastornos depresivos excluidos)

Positivo si la puntuación directa de CC es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad	Positivo si la puntuación tasa base de CC es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad
-1.00	1.000	1.000	-1.00	1.000	1.000
0.50	1.000	0.978	1.50	1.000	0.975
1.50	1.000	0.913	3.50	1.000	0.966
2.50	1.000	0.880	4.50	1.000	0.958
3.50	0.944	0.815	5.50	1.000	0.924
4.50	0.944	0.783	7.00	1.000	0.916
5.50	0.944	0.728	8.50	1.000	0.908
6.50	0.944	0.696	9.50	1.000	0.899
7.50	0.944	0.674	11.50	0.952	0.874
9.00	0.944	0.641	14.00	0.952	0.832
10.50	0.833	0.609	15.50	0.905	0.832
11.50	0.778	0.576	16.50	0.905	0.815
13.00	0.778	0.565	17.50	0.905	0.807
14.50	0.778	0.511	18.50	0.905	0.798
15.50	0.778	0.457	20.50	0.905	0.782
16.50	0.778	0.413	24.00	0.905	0.723
17.50	0.778	0.380	28.00	0.857	0.697
18.50	0.722	0.348	30.50	0.810	0.647
19.50	0.667	0.337	32.50	0.810	0.630
20.50	0.611	0.326	35.00	0.810	0.622
21.50	0.556	0.304	36.50	0.762	0.613
22.50	0.500	0.250	37.50	0.762	0.605
23.50	0.500	0.228	38.50	0.762	0.588
25.00	0.500	0.217	39.50	0.762	0.571
26.50	0.500	0.196	40.50	0.762	0.546

Positivo si la puntuación directa de CC es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad	Positivo si la puntuación base de CC es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad
27.50	0.500	0.185	41.50	0.762	0.538
29.00	0.500	0.174	42.50	0.762	0.504
30.50	0.500	0.152	44.00	0.762	0.496
31.50	0.444	0.141	45.50	0.762	0.445
32.50	0.444	0.130	46.50	0.762	0.437
33.50	0.333	0.120	47.50	0.762	0.403
34.50	0.333	0.109	48.50	0.714	0.395
35.50	0.278	0.087	50.00	0.714	0.387
36.50	0.278	0.076	51.50	0.714	0.370
37.50	0.167	0.065	52.50	0.667	0.353
38.50	0.111	0.054	53.50	0.667	0.345
39.50	0.111	0.043	54.50	0.667	0.336
40.50	0.056	0.033	55.50	0.667	0.328
43.00	0.056	0.022	57.00	0.619	0.311
46.50	0.000	0.011	58.50	0.571	0.303
49.00	0.000	0.000	59.50	0.571	0.294
			60.50	0.524	0.294
			61.50	0.524	0.286
			62.50	0.524	0.277
			63.50	0.524	0.252
			64.50	0.524	0.244
			65.50	0.524	0.235
			67.00	0.476	0.235
			68.50	0.476	0.227
			69.50	0.476	0.218
			70.50	0.476	0.210
			71.50	0.476	0.193
			72.50	0.476	0.168
			74.00	0.476	0.160
			75.50	0.429	0.151
			77.50	0.429	0.143
			79.50	0.381	0.143
			80.50	0.381	0.134
			81.50	0.381	0.126
			83.00	0.381	0.118
			84.50	0.381	0.109
			86.00	0.333	0.109
			88.00	0.333	0.101
			89.50	0.286	0.084
			90.50	0.238	0.084
			91.50	0.238	0.076
			93.50	0.238	0.059
			95.50	0.190	0.059

Positivo si la puntuación directa de CC es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad	Positivo si la puntuación tasa base de CC es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad
			96.50	0.143	0.042
			100.00	0.143	0.025
			104.00	0.095	0.025
			105.50	0.048	0.017
			107.50	0.048	0.008
			109.50	0.000	0.008
			111.00	0.000	0.000

*Nota.* <sup>a</sup>El punto de corte menor es la mínima puntuación observada menos 1, mientras que el punto de corte mayor es la máxima puntuación observada más 1. Todos los demás puntos de corte son la media de dos puntuaciones observadas ordenadas y consecutivas. CC = Escala de Depresión Mayor del MCMI-II.

En la Tabla 6.25 se puede observar que, con respecto a la escala de D del MCMI-II, sensibilidades superiores a 0.70 para el diagnóstico del trastorno depresivo mayor se alcanzaban con puntuaciones directas y TB inferiores a 28 y 66, respectivamente, mientras que especificidades superiores a 0.70 suponían puntuaciones directas y TB superiores a 30 y 70, respectivamente.

**Tabla 6.25.** Coordenadas de las curvas ROC de las puntuaciones directas y de tasa base (TB) de la escala de Distimia (D) del MCMI-II con respecto al diagnóstico de trastorno depresivo mayor (pacientes con el diagnóstico de otros trastornos depresivos excluidos)

Positivo si la puntuación directa de D es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad	Positivo si la puntuación tasa base de D es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad
-1.00	1.000	1.000	-1.00	1.000	1.000
0.50	1.000	0.989	4.00	1.000	0.992
1.50	1.000	0.978	10.50	1.000	0.983
2.50	1.000	0.946	14.50	1.000	0.975
3.50	1.000	0.924	16.50	1.000	0.966
4.50	1.000	0.880	18.00	1.000	0.958
5.50	1.000	0.837	19.50	1.000	0.941
6.50	1.000	0.793	21.00	1.000	0.916
7.50	1.000	0.739	22.50	1.000	0.899
8.50	0.947	0.717	23.50	1.000	0.882

Positivo si la puntuación directa de D es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad	Positivo si la puntuación tasa base de D es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad
9.50	0.947	0.696	24.50	1.000	0.857
10.50	0.947	0.674	26.00	1.000	0.849
11.50	0.895	0.663	27.50	1.000	0.840
12.50	0.895	0.641	28.50	1.000	0.832
13.50	0.895	0.598	29.50	0.955	0.798
14.50	0.895	0.576	30.50	0.955	0.782
15.50	0.895	0.554	32.00	0.955	0.773
16.50	0.895	0.543	33.50	0.955	0.731
17.50	0.895	0.533	34.50	0.955	0.714
18.50	0.842	0.522	35.50	0.955	0.706
19.50	0.842	0.511	37.00	0.955	0.697
20.50	0.842	0.457	38.50	0.909	0.681
21.50	0.842	0.424	39.50	0.909	0.672
22.50	0.842	0.391	40.50	0.909	0.664
23.50	0.842	0.380	41.50	0.909	0.639
24.50	0.842	0.370	42.50	0.909	0.613
25.50	0.789	0.337	43.50	0.909	0.605
26.50	0.789	0.326	44.50	0.909	0.571
27.50	0.737	0.326	45.50	0.909	0.563
28.50	0.684	0.315	46.50	0.909	0.546
29.50	0.632	0.315	47.50	0.818	0.546
30.50	0.632	0.293	48.50	0.818	0.538
31.50	0.579	0.272	49.50	0.818	0.521
32.50	0.579	0.250	50.50	0.818	0.496
33.50	0.579	0.239	51.50	0.818	0.479
35.00	0.579	0.228	52.50	0.773	0.479
37.00	0.526	0.217	54.00	0.773	0.471
39.00	0.526	0.196	56.00	0.727	0.462
41.50	0.526	0.185	57.50	0.727	0.445
43.50	0.526	0.152	58.50	0.727	0.420
44.50	0.474	0.141	59.50	0.727	0.412
45.50	0.421	0.141	61.00	0.727	0.403
47.00	0.316	0.120	62.50	0.727	0.395
49.00	0.316	0.087	63.50	0.727	0.378
50.50	0.263	0.054	64.50	0.727	0.353
51.50	0.211	0.054	66.00	0.682	0.336
53.00	0.105	0.043	68.00	0.682	0.311
54.50	0.053	0.043	69.50	0.682	0.303
55.50	0.053	0.022	70.50	0.682	0.294
58.00	0.053	0.011	71.50	0.682	0.286
61.00	0.000	0.000	72.50	0.682	0.277
			73.50	0.636	0.277
			74.50	0.636	0.269

Positivo si la puntuación directa de D es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad	Positivo si la puntuación base de D es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad
			75.50	0.636	0.261
			76.50	0.591	0.261
			77.50	0.591	0.252
			78.50	0.545	0.252
			80.00	0.545	0.244
			81.50	0.545	0.235
			82.50	0.545	0.227
			83.50	0.545	0.202
			84.50	0.545	0.185
			85.50	0.545	0.168
			87.00	0.500	0.168
			88.50	0.409	0.160
			90.00	0.409	0.151
			92.00	0.364	0.143
			93.50	0.364	0.134
			94.50	0.318	0.118
			95.50	0.318	0.101
			96.50	0.318	0.084
			97.50	0.273	0.084
			98.50	0.273	0.076
			100.50	0.227	0.067
			103.00	0.227	0.059
			104.50	0.182	0.059
			105.50	0.182	0.042
			106.50	0.136	0.034
			107.50	0.136	0.017
			108.50	0.091	0.017
			111.00	0.045	0.008
			118.50	0.045	0.000
			125.00	0.000	0.000

*Nota.* <sup>a</sup>El punto de corte menor es la mínima puntuación observada menos 1, mientras que el punto de corte mayor es la máxima puntuación observada más 1. Todos los demás puntos de corte son la media de dos puntuaciones observadas ordenadas y consecutivas. D MCMI-II = Escala de Distimia del MCMI-II.

En la Tabla 6.26 se puede observar que, con respecto al BDI-II, sensibilidades para el diagnóstico del trastorno depresivo mayor superiores a 0.70 se alcanzaban con puntuaciones inferiores a 23, mientras que especificidades superiores a 0.70 suponían puntuaciones directas superiores a 23.

**Tabla 6.26.** Coordenadas de las curvas ROC de las puntuaciones directas del BDI-II con respecto al diagnóstico de trastorno depresivo mayor (pacientes con el diagnóstico de otros trastornos depresivos excluidos)

Positivo si la puntuación del BDI-II es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad
-1.00	1.000	1.000
0.50	1.000	0.984
1.50	1.000	0.976
2.50	1.000	0.967
3.50	1.000	0.951
4.50	1.000	0.927
5.50	1.000	0.886
6.50	1.000	0.862
7.50	1.000	0.829
8.50	0.958	0.805
9.50	0.958	0.748
10.50	0.958	0.732
11.50	0.958	0.683
12.50	0.958	0.642
13.50	0.917	0.593
14.50	0.875	0.577
15.50	0.833	0.528
16.50	0.833	0.496
17.50	0.750	0.447
18.50	0.750	0.407
19.50	0.750	0.366
20.50	0.750	0.358
21.50	0.708	0.325
22.50	0.708	0.309
23.50	0.667	0.293
24.50	0.667	0.285
25.50	0.667	0.268
26.50	0.625	0.228
27.50	0.583	0.220
28.50	0.542	0.203
29.50	0.458	0.171
30.50	0.375	0.146
31.50	0.292	0.114
32.50	0.208	0.089
34.00	0.167	0.081
36.00	0.125	0.081
37.50	0.083	0.057
38.50	0.083	0.049
39.50	0.042	0.049
40.50	0.042	0.041



<b>Positivo si la puntuación del BDI-II es mayor o igual que<sup>a</sup></b>	<b>Sensibilidad</b>	<b>1 - especificidad</b>
42.00	0.042	0.033
44.00	0.042	0.008
47.50	0.000	0.008
51.00	0.000	0.000

*Nota.* <sup>a</sup>El punto de corte menor es la mínima puntuación observada menos 1, mientras que el punto de corte mayor es la máxima puntuación observada más 1. Todos los demás puntos de corte son la media de dos puntuaciones observadas ordenadas y consecutivas.

En las Tablas 6.27, 6.28 y 6.29, se recogen más índices de validez diagnóstica en relación con el diagnóstico del trastorno depresivo mayor frente a otro trastorno mental no depresivo, para las puntuaciones TB 75 y TB 85 de las escalas CC y D del MCMI-II, para las puntuaciones 14, 20 y 28 del BDI-II y para algunas puntuaciones cercanas a las anteriores.

Como se puede comprobar en la Tabla 6.27, los puntos de corte propuestos por los autores del MCMI-III para la escala CC (TB 75 y TB 85) obtuvieron el segundo valor más elevado de concordancia con el diagnóstico de referencia ( $\kappa = 0.27$ , para ambos puntos de corte). Por otro lado, la puntuación TB 75 tuvo el índice de Youden más elevado (31.65), con una sensibilidad de 47.62% y una especificidad de 84.03%. Entre las puntuaciones directas, la puntuación directa 18 de la escala de CC obtuvo un índice de Youden de 39.74, con una sensibilidad de 77.78%, una especificidad de 61.96% y un índice  $\kappa$  de 0.23, mientras que la puntuación directa 32 obtuvo el índice  $\kappa$  más elevado, de 0.28, con un índice de Youden de 30.31 y una sensibilidad menor que la especificidad (44.44% y 85.87%, respectivamente). En conjunto, los mejores puntos de corte para la escala CC obtuvieron índices  $\kappa$  en los valores bajos del rango que indica una concordancia débil (0.21-0.40; Altman, 1991), y aunque ninguno obtuvo un equilibrio adecuado entre sensibilidad y especificidad, al menos mostraron o bien una

especificidad adecuada, es decir, > 70% (TB 75, TB 85 y la puntuación directa 32) o bien una sensibilidad adecuada, es decir, > 70% (puntuación directa 18). Por otro lado, en todos los puntos de corte estudiados para la escala CC, el poder predictivo positivo fue débil (< 34%) y el poder predictivo positivo fue alto (> 88%).

En el caso de la escala D del MCMI-II (véase Tabla 6.28), la puntuación con un índice kappa más elevado fue la puntuación TB 85 (0.3), con una sensibilidad de 54.55% y una especificidad de 81.51%, de manera que el índice de Youden fue 36.06. La puntuación TB 75 obtuvo una especificidad menor (63.64%) y una sensibilidad superior (73.11%), pero unos índices kappa y de Youden ligeramente inferiores de 0.25 y 36.75, respectivamente. La puntuación TB con un mayor índice de Youden (37.09) fue la TB 68, con una mayor sensibilidad que especificidad (68.18% y 68.91%, respectivamente) y un índice kappa de 0.24. Un índice de Youden más alto (45.08) fue obtenido por la puntuación directa de 23, la cual tuvo, además, un índice kappa de 0.27, una sensibilidad de 84.21% y una especificidad de 60.87%. La puntuación directa con un índice kappa más elevado fue la puntuación 35 (0.28), con una especificidad de 77.17% y una sensibilidad de 58.89%. Por otro lado, tal y como ocurría con la escala CC, los mejores puntos de corte para la escala D obtuvieron índices kappa en los valores bajos del rango que indica una concordancia débil (0.21-0.40; Altman, 1991), y aunque ninguno obtuvo un equilibrio adecuado entre sensibilidad y especificidad, al menos mostraron o bien una especificidad adecuada, es decir, > 70% (TB 75, TB 85 y puntuación directa 35) o bien una sensibilidad adecuada, es decir, > 70% (puntuación directa 23); además, para todos los puntos de corte de la escala D estudiados, el poder predictivo positivo fue débil (< 45%) y el poder predictivo positivo fue alto (> 89%).

En el caso del BDI-II (véase la Tabla 6.29), la puntuación 27 presentó el mayor índice kappa (0.3), con un índice de Youden de 39.74, una especificidad de 77.24% y una

sensibilidad de 62.5%. Otras puntuaciones del BDI-II con niveles más elevados de sensibilidad como la puntuación 23 (70.83%), obtuvo un índice kappa (0.26) y una especificidad (69.11%) inferiores, aunque un índice de Youden ligeramente superior (39.94). No obstante, los índices kappa también se situaban en el rango bajo de los estándares para una concordancia débil (0.21-0.40) y ninguna puntuación obtuvo un adecuado equilibrio entre sensibilidad y especificidad, aunque había puntuaciones con sensibilidades adecuadas (puntuación de 23) y especificidades adecuadas (puntuación de 27).

**Tabla 6.27.** Índices de validez diagnóstica de distintas puntuaciones directas (PD) o tasa base (TB) de la escala de Depresión Mayor (CC) del MCMI-II con respecto al diagnóstico del trastorno depresivo (pacientes con el diagnóstico de otros trastornos depresivos excluidos)

Punto de corte – Escala	Sensibilidad	Especificidad	Índice de Youden	Poder predictivo positivo	Poder predictivo negativo	Tasa de falsos positivos	Tasa de falsos negativos	Kappa	Eficacia
<b>TB 75</b>	<b>47.62</b>	<b>84.03</b>	<b>31.65</b>	<b>34.48</b>	<b>90.09</b>	<b>15.97</b>	<b>52.38</b>	<b>0.27</b>	<b>78.57</b>
<b>TB 85</b>	<b>38.1</b>	<b>89.08</b>	<b>27.18</b>	<b>38.1</b>	<b>89.08</b>	<b>10.92</b>	<b>61.9</b>	<b>0.27</b>	<b>81.43</b>
TB 66	52.38	76.47	28.85	28.21	90.1	23.53	47.62	0.21	72.86
TB 67	47.62	76.47	24.09	26.32	89.22	23.53	52.38	0.18	72.14
TB 68	47.62	76.47	24.09	26.32	89.22	23.53	52.38	0.18	72.14
TB 62	52.38	71.43	23.81	24.44	89.47	28.57	47.62	0.16	68.57
TB 61	52.38	70.59	22.97	23.91	89.36	29.41	47.62	0.15	67.86
TB 60	57.14	70.59	27.73	25.53	90.32	29.41	42.86	0.18	68.57
PD 7	94.44	30.43	24.87	20.99	96.55	69.57	5.56	0.1	40.91
PD 9	94.44	35.87	30.31	22.37	97.06	64.13	5.56	0.13	45.45
PD 8	94.44	32.61	27.05	21.52	96.77	67.39	5.56	0.11	42.73
PD 10	94.44	35.87	30.31	22.37	97.06	64.13	5.56	0.13	45.45
PD 11	83.33	40.45	23.78	22.06	92.31	59.55	16.67	0.11	47.66
PD 12	77.78	42.39	20.17	20.9	90.7	57.61	22.22	0.1	48.18
PD 13	77.78	43.48	21.26	21.21	90.91	56.52	22.22	0.1	49.09
PD 14	77.78	43.48	21.26	21.21	90.91	56.52	22.22	0.1	49.09
PD 15	77.78	48.91	26.69	22.95	91.84	51.09	22.22	0.14	53.64
PD 16	77.78	54.35	32.13	25	92.59	45.65	22.22	0.17	58.18
PD 17	77.78	58.7	36.48	26.92	93.1	41.3	22.22	0.21	61.82
<b>PD 18</b>	<b>77.78</b>	<b>61.96</b>	<b>39.74</b>	<b>28.57</b>	<b>93.44</b>	<b>38.04</b>	<b>22.22</b>	<b>0.23</b>	<b>64.55</b>
PD 19	72.22	65.22	37.44	28.89	92.31	34.78	27.78	0.23	66.36
PD 20	66.67	66.3	32.97	27.91	91.04	33.7	33.33	0.21	66.36
PD 23	50	75	25	28.13	88.46	25	50	0.19	70.91
PD 28	50	81.52	31.52	34.62	89.29	18.48	50	0.27	76.36
<b>PD 32</b>	<b>44.44</b>	<b>85.87</b>	<b>30.31</b>	<b>38.1</b>	<b>88.76</b>	<b>14.13</b>	<b>55.56</b>	<b>0.28</b>	<b>79.09</b>
PD 35	33.33	89.13	22.46	37.5	87.23	10.87	66.67	0.24	80

**Tabla 6.28.** Índices de validez diagnóstica de distintas puntuaciones directas (PD) o tasa base (TB) de la escala de Distimia (D) del MCMI-II con respecto al diagnóstico del trastorno depresivo mayor (pacientes con el diagnóstico de otros trastornos depresivos excluidos)

Punto de corte – Escala	Sensibilidad	Especificidad	Índice de Youden	Poder predictivo positivo	Poder predictivo negativo	Tasa de falsos positivos	Tasa de falsos negativos	Kappa	Eficiencia
<b>TB 85</b>	<b>54.55</b>	<b>81.51</b>	<b>36.06</b>	<b>35.29</b>	<b>90.65</b>	<b>18.49</b>	<b>45.45</b>	<b>0.3</b>	<b>77.3</b>
<b>TB 75</b>	<b>63.64</b>	<b>73.11</b>	<b>36.75</b>	<b>30.43</b>	<b>91.58</b>	<b>26.89</b>	<b>36.36</b>	<b>0.25</b>	<b>71.63</b>
TB 66	68.18	66.39	34.57	27.27	91.86	33.61	31.82	0.21	66.67
TB 67	68.18	66.39	34.57	27.27	91.86	33.61	31.82	0.21	66.67
<b>TB 68</b>	<b>68.18</b>	<b>68.91</b>	<b>37.09</b>	<b>28.85</b>	<b>92.13</b>	<b>31.09</b>	<b>31.82</b>	<b>0.24</b>	<b>68.79</b>
TB 60	72.73	58.82	31.55	24.62	92.11	41.18	27.27	0.18	60.99
TB 61	72.73	59.66	32.39	25	92.21	40.34	27.27	0.18	61.7
TB 62	72.73	59.66	32.39	25	92.21	40.34	27.27	0.18	61.7
PD 8	100	26.09	26.09	21.84	100	73.91	0	0.11	38.74
PD 9	94.74	28.26	23	21.43	96.3	71.74	5.26	0.1	39.64
PD 10	94.74	30.43	25.17	21.95	96.55	69.57	5.26	0.11	41.44
PD 11	94.74	32.61	27.35	22.5	96.77	67.39	5.26	0.12	43.24
PD 15	89.47	42.39	31.86	24.29	95.12	57.61	10.53	0.15	50.45
PD 18	89.47	46.74	36.21	25.76	95.56	53.26	10.53	0.18	54.05
PD 20	84.21	48.91	33.12	25.4	93.75	51.09	15.79	0.17	54.95
<b>PD 23</b>	<b>84.21</b>	<b>60.87</b>	<b>45.08</b>	<b>30.77</b>	<b>94.92</b>	<b>39.13</b>	<b>15.79</b>	<b>0.27</b>	<b>64.86</b>
PD 28	73.68	67.39	41.07	31.82	92.54	32.61	26.32	0.27	68.47
PD 32	57.89	72.83	30.72	30.56	89.33	27.17	42.11	0.23	70.27
<b>PD 35</b>	<b>57.89</b>	<b>77.17</b>	<b>35.06</b>	<b>34.38</b>	<b>89.87</b>	<b>22.83</b>	<b>42.11</b>	<b>0.28</b>	<b>73.87</b>

**Tabla 6.29.** Índices de validez diagnóstica de distintas puntuaciones del BDI-II con respecto al diagnóstico de trastorno depresivo mayor (pacientes con el diagnóstico de otros trastornos depresivos excluidos)

Punto de corte — Escala	Sensibilidad	Especificidad	Índice de Youden	Poder predictivo positivo	Poder predictivo negativo	Tasa de falsos positivos	Tasa de falsos negativos	Kappa	Eficiencia
<b>BDI-II 14</b>	91,67	40,65	32,32	23,16	96,15	59,35	8,33	0,15	48,98
BDI-II 15	87,5	42,28	29,78	22,83	94,55	57,72	12,5	0,14	49,66
BDI-II 16	83,33	47,15	30,48	23,53	93,55	52,85	16,67	0,15	53,06
BDI-II 17	83,33	50,41	33,74	24,69	93,94	49,59	16,67	0,17	55,78
BDI-II 18	75	55,28	30,28	24,66	91,89	44,72	25	0,17	58,5
BDI-II 19	75	59,35	34,35	26,47	92,41	40,65	25	0,2	61,9
BDI-II 20	75	63,41	38,41	28,57	92,86	36,59	25	0,23	65,31
BDI-II 21	75	64,23	39,23	29,03	92,94	35,77	25	0,24	65,99
BDI-II 22	70,83	67,48	38,31	29,82	92,22	32,52	29,17	0,25	68,03
<b>BDI-II 23</b>	<b>70,83</b>	<b>69,11</b>	<b>39,94</b>	<b>30,91</b>	<b>92,39</b>	<b>30,89</b>	<b>29,17</b>	<b>0,26</b>	<b>69,39</b>
BDI-II 24	66,67	70,73	37,4	30,77	91,58	29,27	33,33	0,25	70,07
BDI-II 25	66,67	71,54	38,21	31,37	91,67	28,46	33,33	0,26	70,75
<b>BDI-II 27</b>	<b>62,5</b>	<b>77,24</b>	<b>39,74</b>	<b>34,88</b>	<b>91,35</b>	<b>22,76</b>	<b>37,5</b>	<b>0,3</b>	<b>74,83</b>
BDI-II 28	58,33	78,05	36,38	34,15	90,57	21,95	41,67	0,28	74,83
BDI-II 29	54,17	79,67	33,84	34,21	89,91	20,33	45,83	0,27	75,51
BDI-II 30	45,83	82,93	28,76	34,38	88,7	17,07	54,17	0,25	76,87
BDI-II 31	37,5	85,37	22,87	33,33	87,5	14,63	62,5	0,22	77,55
BDI-II 32	29,17	88,62	17,79	33,33	86,51	11,38	70,83	0,19	78,91
BDI-II 33	20,83	91,06	11,89	31,25	85,5	8,94	79,17	0,14	79,59
BDI-II 34	16,67	91,87	8,54	28,57	84,96	8,13	83,33	0,1	79,59
BDI-II 35	16,67	91,87	8,54	28,57	84,96	8,13	83,33	0,1	79,59

#### 6.4. Discusión

El objetivo general del presente estudio era analizar la validez diagnóstica de las escalas de Depresión Mayor (CC) y Distimia (D) de la adaptación española del MCMI-II, ya que, desde su publicación por primera vez en España (Millon, 1998, adaptación española de A. Ávila-Espada), sus usuarios las están utilizando sin conocer exactamente sus propiedades diagnósticas, precisamente uno de los principales usos para los que fueron diseñadas (Millon, 1998). Además, hasta ahora, a nivel internacional, tan sólo se han publicado tres estudios con datos empíricos sobre la validez diagnóstica de las escalas CC y D del MCMI-II (Millon, 2002; Piersma, 1991; Wetzler y Marlowe, 1993), por lo que el presente estudio parecía necesario en el contexto de una literatura científica empírica tan escasa y que no hace justicia a la gran popularidad del instrumento tanto en España como en otros países.

En concreto, a partir de una muestra ambulatoria de pacientes con diversos trastornos mentales, este estudio pretendía, en primer lugar, examinar la validez diagnóstica de las puntuaciones TB 75 y TB 85 (las puntuaciones de corte estándar o habituales en el MCMI-II) de las escalas CC y D para el diagnóstico del trastorno depresivo mayor y para el diagnóstico del conjunto de trastornos depresivos. En segundo lugar, el estudio pretendía conocer si existían otras puntuaciones de las escalas CC y D del MCMI-II que tuvieran mejor validez diagnóstica que las puntuaciones de corte habituales de dichas escalas. Finalmente, el estudio también intentaba examinar si los índices de validez diagnóstica de las escalas CC y D del MCMI-II eran mejores que los de otros instrumentos para evaluar la depresión ampliamente utilizados, en concreto, mejores que el BDI-II.

**¿Cuál es la validez diagnóstica de las puntuaciones TB 75 y TB 85 de las escalas CC y D del MCMI-II para el diagnóstico del trastorno depresivo mayor?**

Respecto al diagnóstico del trastorno depresivo mayor, los resultados obtenidos sugieren, en general, que la validez diagnóstica de ambas escalas es cuestionable, ya que la mayoría de sus índices de rendimiento diagnóstico se encontraban por debajo de los estándares aceptables para un instrumento diagnóstico. Efectivamente, para ambas puntuaciones y para ambas escalas, los índices de sensibilidad se encontraban por debajo del estándar de 70%, los índices de concordancia diagnóstica (kappa) estaban por debajo del estándar de .40, y los índices de Youden no reflejaban un equilibrio adecuado entre sensibilidad y especificidad ( $< 40$ ). Tan sólo los índices de especificidad apoyaban la validez diagnóstica de las escalas CC y D del MCMI-II para el diagnóstico del trastorno depresivo mayor, ya que, para las dos puntuaciones de corte (TB 75 y TB 85) y para las dos escalas, todas las especificidades superaban el estándar considerado aceptable ( $> 70\%$ ).

Estos índices de especificidad aceptables indicarían que las escalas CC y D del MCMI-II son útiles para detectar correctamente a un paciente que no tiene el trastorno depresivo mayor. De hecho, y de manera relacionada, para las dos puntuaciones de corte y para las dos escalas, sus índices de poder o valor predictivo negativo eran muy buenos y superaban el 89%. El poder o valor predictivo negativo es un índice de rendimiento diagnóstico muy útil en la práctica clínica, ya que indica, en el caso de que un paciente puntúe por debajo de la TB 75 o la TB 85 en la escala CC o D del MCMI-II, la probabilidad de que ese paciente no tenga efectivamente un trastorno depresivo mayor. Es decir, sobre la base de los resultados de la presente muestra de pacientes, se podría estimar que los pacientes que puntúen por debajo de la TB 75 o de la TB 85 en la escala



CC o D del MCMI-II, tienen más del 89% de probabilidades de no tener un trastorno depresivo mayor.

Sin embargo, tan importantes como la especificidad y el poder o valor predictivo negativo de un instrumento, lo son su sensibilidad y su poder o valor predictivo positivo. En concreto, el poder predictivo positivo es un índice también muy útil desde el punto de vista clínico, ya que indica, en el caso de que un paciente puntúe por encima de la TB 75 o de la TB 85 en la escala CC o D del MCMI-II, la probabilidad de que ese paciente tenga efectivamente un trastorno depresivo mayor. En este sentido, los valores predictivos positivos de las puntuaciones TB 75 y TB 85 de las escalas CC y D del MCMI-II se situaban todos por debajo del 38%, es decir, que la probabilidad de que un paciente que tiene una puntuación de TB 75 o TB 85 en las escalas CC o D padezca en realidad un trastorno depresivo mayor es menor del 50%, o dicho de otra manera, menor que la probabilidad que se obtendría de arrojar una moneda al aire para decidir si padece o no un trastorno depresivo mayor.

Este patrón de resultados débiles sobre el rendimiento diagnóstico de las puntuaciones de corte habituales de las escalas CC y D del MCMI-II fue también evidente incluso cuando, siguiendo las recomendaciones de Ganellen (1996), se evaluó la eficacia de las escalas para diagnosticar un trastorno depresivo mayor frente a otro trastorno mental no depresivo, es decir, excluyendo de la muestra de pacientes del presente estudio los pacientes que tenían otros diagnósticos de depresión distintos al trastorno depresivo mayor (distimia, trastorno depresivo no especificado, duelo y trastorno adaptativo con estado de ánimo depresivo o mixto). De hecho, los índices de validez diagnóstica fueron en todos los casos muy similares (véase la Tabla 6.30).

Como puede observarse en la Tabla 6.30, todos estos resultados sobre el rendimiento de las escalas CC y D del MCMI-II para diagnosticar un trastorno depresivo

mayor difieren de los obtenidos en los dos estudios previos publicados hasta la fecha (Millon, 2002; Piersma, 1991), aunque en direcciones distintas.

**Tabla 6.30.** Índices de validez diagnóstica de las escalas de Depresión Mayor (CC) y Distimia (D) del MCMI-II para el diagnóstico del trastorno depresivo mayor en muestras heterogéneas de pacientes con trastornos mentales

Índices	Millon (2002)		Piersma (1991) <sup>a</sup>				Este estudio				Este estudio <sup>b</sup>			
	CC		CC		D		CC		D		CC		D	
	TB 75	TB 85	TB 75	TB 85	TB 75	TB 85	TB 75	TB 85	TB 75	TB 85	TB 75	TB 85	TB 75	TB 85
Prev.	14	6	77	ni	77	ni	13	13	13	13	16	16	16	16
Sens.	68	47	61	ni	86	ni	47	38	63	54	47	38	63	54
Espec.	96	97	52	ni	32	ni	83	88	73	82	89	84	81	73
VPP	73	50	81	ni	81	ni	27	30	26	30	34	38	30	35
VPN	95	97	28	ni	40	ni	92	91	93	92	90	89	91	90
PDT	92	94	59	ni	73	ni	79	81	72	78	78	81	71	77
Kappa	ni	ni	ni	ni	ni	ni	.24	.24	.23	.27	.27	.27	.25	.3
Youden	64 <sup>c</sup>	44 <sup>c</sup>	13 <sup>c</sup>	ni	18 <sup>c</sup>	ni	31	26	37	36	31	27	36	36

*Nota.* Prev.: prevalencia. Sens. = sensibilidad. Espec. = especificidad. VPP: valor o poder predictivo positivo. VPN: valor o poder predictivo negativo. PDT: poder diagnóstico total o índice de eficacia global (porcentaje de pacientes correctamente clasificados) (también denominado precisión, índice de validez o de eficiencia, o valor diagnóstico total). ni: no informado.

<sup>a</sup>La muestra de este estudio no era totalmente heterogénea, ya que todos los pacientes tenían trastornos depresivos (trastorno depresivo mayor, distímico, adaptativo con estado de ánimo deprimido y depresivo no especificado).

<sup>b</sup>La muestra no era totalmente heterogénea, ya que estaban excluidos los pacientes con otros trastornos depresivos distintos al trastorno depresivo mayor (la muestra incluía pacientes con trastorno depresivo mayor y pacientes con otros trastornos mentales no depresivos).

<sup>c</sup>Este índice no aparece recogido en la publicación correspondiente, sino que ha sido calculado para este estudio a partir de los datos de la tabla.

Como cabría esperar, los resultados de rendimiento diagnóstico obtenidos por el equipo de investigación que desarrolló el MCMI-II son más positivos (Millon, 2002). En concreto, en el manual del MCMI-II (Millon, 2002) se proporcionaban índices adecuados para la escala CC a la hora de diagnosticar un trastorno depresivo mayor, con una alta especificidad para las puntuaciones TB 75 y TB 85 (96% y 97%, respectivamente) y con

una sensibilidad que casi rozaba el estándar de 70% para la puntuación TB 75 (68%), lo cual permitía obtener poderes predictivos negativos excelentes ( $> 95\%$ ) para ambas puntuaciones de corte, pero también un poder predictivo positivo mayor de 50% para la puntuación TB 75 (73%).

En resumen, como se puede apreciar al comparar los resultados de Millon (2002) y los del presente estudio (véase la Tabla 6.30), aunque en ambos estudios las escalas CC y D del MCMI-II, con sus puntuaciones de corte habituales (TB 75 y TB 85), se mostraban como instrumentos adecuados para descartar la presencia de un trastorno depresivo mayor (tal y como refleja la magnitud de los índices de especificidad y valor predictivo negativo en ambos estudios), en el presente estudio dichas escalas no fueron lo suficientemente sensibles para detectar a los pacientes con trastorno depresivo mayor, al contrario de lo que ocurrió en el estudio original de Millon (2002), en el cual, al menos para la puntuación TB 75, se obtuvieron índices aceptables o casi aceptables de sensibilidad y de valor predictivo positivo, los cuales revertían en un mayor equilibrio entre sensibilidad y especificidad (índices de Youden  $> 40$ ).

Lamentablemente, Millon (2002) no informaba de otros índices de rendimiento diagnóstico también importantes para valorar la validez diagnóstica de un instrumento como, por ejemplo, el índice kappa de concordancia con el diagnóstico de referencia o el área bajo la curva ROC, los cuales sí fueron obtenidos en el presente estudio. Sobre este problema se volverá más adelante, cuando se comente el área bajo la curva ROC como un índice global del rendimiento diagnóstico de un instrumento. Baste señalar por ahora que otros investigadores también han llamado la atención sobre la necesidad de conocer una mayor amplitud de índices diagnósticos como, por ejemplo, el índice kappa, para establecer la validez diagnóstica de un instrumento (Hsu, 2002).

Por otro lado, cuando se comparan los resultados del presente estudio con los resultados encontrados por Piersma (1991), ambos coinciden en sugerir la débil eficacia diagnóstica de las escalas CC y D del MCMI-II para diagnosticar un trastorno depresivo mayor, pero por motivos distintos. Como se puede apreciar en la Tabla 6.30, en el estudio de Piersma (1991), y al contrario de lo que ocurría en el presente estudio y en el estudio de Millon (2002), los índices de sensibilidad y valor predictivo positivo fueron mejores que los índices de especificidad y valor predictivo negativo. Esta diferencia se podría explicar por el hecho de que la muestra evaluada en el estudio de Piersma (1991) no era tan heterogénea como la evaluada en el presente estudio o en el estudio de Millon (2002), ya que todos los pacientes de Piersma (1991) tenían trastornos depresivos (trastorno depresivo mayor, distímico, adaptativo con estado de ánimo deprimido y depresivo no especificado) y, además, entre estos trastornos depresivos, el trastorno depresivo mayor era el más frecuente con mucha diferencia, de manera que la prevalencia de este trastorno en la muestra de pacientes de Piersma (1991) era muchísimo más alta que en el presente estudio o en el estudio de Millon (2002) (77% frente a 6-14%; véase la Tabla 6.30). De hecho, estas discrepancias “ilustran cómo los procedimientos de selección [de las muestras de pacientes] que los investigadores adoptan pueden afectar a los índices de eficiencia diagnóstica” (Ganellen, 1996, p. 229).

Más allá de esas discrepancias, los resultados del estudio de Piersma (1991) coincidían con los del presente estudio en cuestionar la validez de las puntuaciones TB 75 de las escalas CC y D del MCMI-II para diagnosticar un trastorno depresivo mayor, ya que, de forma coincidente con el presente estudio, pero de forma contraria al estudio de Millon (2002), la mayoría de sus índices de rendimiento diagnóstico (3 sobre 5) se encontraban por debajo de los estándares aceptables para un instrumento diagnóstico. Así, en el estudio de Piersma (1991), mientras que los índices de sensibilidad y valor predictivo positivo se encontraban por encima del estándar (> 70%), los índices de

especificidad y valor predictivo negativo se situaban muy por debajo del estándar de 70% (32-52% y 28-40%, respectivamente), lo que daba lugar a unos índices de Youden bajos (13-18%) que reflejaban importantes desequilibrios entre sensibilidad y especificidad.

**¿Cuál es la validez diagnóstica de las puntuaciones TB 75 y TB 85 de las escalas CC y D del MCMI-II para el diagnóstico de cualquier trastorno depresivo?**

Respecto al diagnóstico del conjunto de trastornos depresivos, los resultados obtenidos en el presente estudio sugieren, en general y de forma coincidente con los obtenidos respecto al diagnóstico del trastorno depresivo mayor, que la validez diagnóstica de ambas escalas para diagnosticar cualquier trastorno depresivo es cuestionable, ya que la mayoría de sus índices de rendimiento diagnóstico se encontraban por debajo de los estándares aceptables para un instrumento diagnóstico. Efectivamente, para ambas puntuaciones y para ambas escalas, los índices de sensibilidad se encontraban por debajo del estándar de 70%, los índices de concordancia diagnóstica (kappa) estaban por debajo del estándar de .40, y los índices de Youden no reflejaban un equilibrio adecuado entre sensibilidad y especificidad ( $< 40$ ). Tan sólo los índices de especificidad apoyaban la validez diagnóstica de las escalas CC y D del MCMI-II para el diagnóstico de cualquier trastorno depresivo, ya que, para las dos puntuaciones de corte (TB 75 y TB 85) y para las dos escalas, todas las especificidades superaban el estándar considerado aceptable ( $> 70\%$ ).

Estos índices de especificidad aceptables, así como los adecuados valores predictivos negativos ( $> 70\%$ ), indicarían que las escalas CC y D del MCMI-II podrían ser útiles para detectar correctamente a un paciente que no tiene un trastorno depresivo; sin embargo, sus índices tan bajos de sensibilidad ( $< 41\%$ ) y de valor predictivo positivo ( $< 50\%$ ) también sugieren que la probabilidad de que un paciente que tiene una puntuación de TB 75 y TB 85 en las escalas CC o D padezca en realidad cualquier trastorno

depresivo es menor que la probabilidad que se obtendría de arrojar una moneda al aire para decidir si padece o no un trastorno depresivo (en el caso de la puntuación TB 85 de la escala de CC, la especificidad es el 50%, de manera que en este caso la probabilidad de acertar sería exactamente la misma que al arrojar una moneda al aire).

Como se puede observar en la Tabla 6.31, estos resultados sobre el rendimiento de las escalas CC y D del MCMI-II para diagnosticar un trastorno depresivo mayor son similares a los obtenidos en el único estudio previo publicado hasta la fecha y que ha abordado también la validez diagnóstica respecto al conjunto de trastornos depresivos (Wetzler y Marlowe, 1993).

Efectivamente, al igual que ocurría en el presente estudio, en el estudio de Wetzler y Marlowe (1993) los análisis de eficacia diagnóstica también revelaron, para ambas puntuaciones y para ambas escalas, sensibilidades que encontraban por debajo del estándar de 70% e índices de Youden que no reflejaban un equilibrio adecuado entre sensibilidad y especificidad ( $< 40$ ) y que, de hecho, en el caso de la escala de Depresión Mayor (CC) eran extremadamente bajos (4 y 2 para las puntuaciones TB 75 y TB 85, respectivamente). Tan sólo los índices de especificidad apoyaban la validez diagnóstica de las escalas CC y D del MCMI-II para el diagnóstico de cualquier trastorno depresivo, ya que, para las dos puntuaciones de corte (TB 75 y TB 85) y para las dos escalas, todas las especificidades superaban el estándar considerado aceptable ( $> 70\%$ ).

**Tabla 6.31.** Índices de validez diagnóstica de las escalas de Depresión Mayor (CC) y Distimia (D) del MCMI-II para el diagnóstico de cualquier trastorno depresivo en muestras heterogéneas de pacientes con trastornos mentales

Índices	Wetzler y Marlowe (1993)				Este estudio			
	Escala de Depresión (CC)		Escala de Distimia (D)		Escala de Depresión (CC)		Escala de Distimia (D)	
	TB 75	TB 85	TB 75	TB 85	TB 75	TB 85	TB 75	TB 85
Prevalencia	55	55	55	55	31	31	31	31
Sensibilidad	18	10	57	33	30	25	32	41

Índices	Wetzler y Marlowe (1993)				Este estudio			
	Escala de Depresión (CC)		Escala de Distimia (D)		Escala de Depresión (CC)		Escala de Distimia (D)	
	TB 75	TB 85	TB 75	TB 85	TB 75	TB 85	TB 75	TB 85
Especificidad	86	92	70	86	84	89	73	81
VPP	47	46	57	52	45	50	43	40
VPN	61	60	69	74	73	73	73	73
PDT	49	47	63	57	68	69	63	66
Kappa	ni	ni	ni	ni	0.16	0.16	0.15	0.15
Youden	4 <sup>a</sup>	2 <sup>a</sup>	27 <sup>a</sup>	19 <sup>a</sup>	15	14	13	13

*Nota.* VPP: valor o poder predictivo positivo. VPN: valor o poder predictivo negativo. PDT: poder diagnóstico total o índice de eficacia global (porcentaje de pacientes correctamente clasificados) (también denominado precisión, índice de validez o de eficiencia, o valor diagnóstico total). ni: no informado.

<sup>a</sup>Este índice no aparece recogido en la publicación correspondiente, sino que ha sido calculado para este estudio a partir de los datos de la tabla.

### ¿Existen otras puntuaciones de las escalas CC y D del MCMI-II que tengan mejores índices de validez diagnóstica que sus puntuaciones de corte habituales (TB 75 y TB 85)?

Los resultados del presente estudio sugieren que las puntuaciones TB 75 y TB 85 de las escalas CC y D del MCMI-II, cuando se utilizan como puntos de corte para el diagnóstico del trastorno depresivo mayor, son, entre todas las puntuaciones TB analizadas de dichas escalas, las que mostraban, en general, los mejores índices de validez diagnóstica (p. ej., índices kappa y de Youden más altos), aunque las diferencias con otras puntuaciones TB fueron muy pequeñas. Sin embargo, entre las puntuaciones directas analizadas, sí que se encontraron no sólo puntos de corte con similares índices de validez diagnóstica, sino incluso algunas puntuaciones directas cuyos índices superaban ligeramente a los de las puntuaciones TB 75 y TB 85. Por ejemplo, en la escala CC, las puntuaciones directas 32 y 18 tenían kappas e índice de Youden ligeramente más elevados para el diagnóstico del trastorno depresivo mayor, mientras que en la escala D, la puntuación directa 23 tenía un índice de Youden ligeramente mayor. En cualquier caso, todas las puntuaciones TB o directas señaladas presentaban,

en general, al igual que les ocurría a las puntuaciones TB 75 y TB 85, índices de validez diagnóstica (sensibilidad, poder predictivo positivo, kappa, índice de Youden) con valores inferiores a los considerados aceptables para un instrumento diagnóstico, con la única salvedad de los índices de especificidad y de poder predictivo negativo. Este patrón de resultados fue prácticamente igual cuando se analizó el rendimiento de diferentes puntuaciones TB y diferentes puntuaciones directas de las escalas CC y D del MCMI-II para diagnosticar un trastorno depresivo mayor frente a un trastorno mental no depresivo (es decir, cuando se eliminó de la muestra de participantes los pacientes con un diagnóstico de otro trastorno depresivo distinto al trastorno depresivo mayor).

Un patrón de resultados también similar fue encontrado cuando se analizó el rendimiento de la escala CC para realizar el diagnóstico de cualquier trastorno depresivo. Las puntuaciones TB 75 y TB 85 fueron de nuevo las puntuaciones TB con índices de validez diagnóstica en general ligeramente mejores, aunque también hubo algunas puntuaciones directas que mostraron un rendimiento diagnóstico ligeramente mejor que las puntuaciones TB 75 y TB 85 (p. ej., las puntuaciones directas 9 y 18). Sin embargo, en el caso de la escala D, las puntuaciones TB 61 y TB 62, al igual que las puntuaciones directas 15 y 23, mostraron índices de validez para el diagnóstico de cualquier trastorno depresivo ligeramente superiores que las puntuaciones TB 75 y TB 85. En cualquier caso, de nuevo todas las puntuaciones TB o directas señaladas presentaban, en general, al igual que les ocurría a las puntuaciones TB 75 y TB 85, índices de validez para el diagnóstico de cualquier trastorno depresivo (sensibilidad, poder predictivo positivo, kappa, índice de Youden) inferiores a los considerados aceptables para un instrumento diagnóstico, con la única salvedad de los índices de especificidad y de poder predictivo negativo.



De estos resultados se podría concluir, en primer lugar, que no está claro que la estrategia utilizada por el MCMI-II de identificar los puntos de corte para el diagnóstico del trastorno depresivo mayor (o de cualquier trastorno depresivo) a partir de las puntuaciones TB, es decir, estableciendo como punto de corte aquella puntuación que deja por encima un porcentaje de la muestra de pacientes que coincide con la prevalencia en dicha muestra del trastorno depresivo mayor (o de los trastornos depresivos en general) (TB 75) o que coincide con la prevalencia en dicha muestra del trastorno depresivo mayor (o de los trastornos depresivos en general) como trastorno más sobresaliente (TB 85), sea siempre la estrategia más adecuada para obtener los puntos de corte con mayor eficacia diagnóstica, ya que se pueden encontrar, a partir de un estudio de rendimiento diagnóstico como el realizado en el presente estudio, otras puntuaciones que obtienen índices al menos similares de eficacia diagnóstica.

De los resultados mencionados anteriormente también se puede extraer una segunda conclusión más importante: no sólo las puntuaciones TB 75 y TB 85 de las escalas CC y D del MCMI-II, sino también otras puntuaciones directas o TB de dichas escalas evaluadas en el presente estudio, presentan, en general, una eficacia diagnóstica cuestionable para diagnosticar un trastorno depresivo mayor o para diagnosticar cualquier trastorno depresivo.

En el caso de los índices de eficiencia diagnóstica con respecto al conjunto de los trastornos depresivos, los resultados son consistentes con los obtenidos mediante el análisis de la curva ROC. De hecho, estos últimos resultados son muy reveladores ya que permiten evaluar la eficacia diagnóstica de un instrumento a lo largo de todo el rango de sus puntuaciones (Swets, 1988). Así, en concreto, las áreas bajo la curva ROC tanto de las puntuaciones TB como de las puntuaciones directas de las escalas CC y D del MCMI-II en el caso del diagnóstico de cualquier trastorno depresivo nunca alcanzaron la

significación estadística, es decir, nunca fueron diferentes de forma estadísticamente significativa del área bajo la curva ROC que representa un diagnóstico basado en el azar. Es más, dichas áreas ni siquiera alcanzaron el valor mínimo que representa una exactitud aceptable ( $> 0.70$ ; Fischer et al., 2003; Streiner y Cairney, 2007; Swets, 1988),

Con respecto al diagnóstico del trastorno depresivo mayor, en algunos casos (las puntuaciones TB y directas de la escala D y las puntuaciones TB de la escala CC) las áreas bajo la curva ROC sí superaron el rango mínimo de las exactitudes aceptables ( $> 0.70$ ) y fueron estadísticamente significativas, pero no alcanzaron exactitudes que otros investigadores más rigurosos consideran necesarias para que tengan significación clínica ( $> 0.75$ ; Fan et al., 2006). En cambio, en el caso de la escala de puntuaciones directas de la escala CC, el área bajo la curva ROC ni siquiera superó el rango mínimo de exactitudes aceptables. Cuando se excluyeron de los análisis los pacientes con otros trastornos depresivos, las áreas de la curva ROC de las puntuaciones TB y directas de las escalas CC y D para el diagnóstico del trastorno depresivo mayor fueron estadísticamente significativas, superaron el rango mínimo de las exactitudes aceptables ( $> 0.70$ ) e incluso, en el caso de las puntuaciones TB y las puntuaciones directas la escala CC, alcanzaron las exactitudes que otros investigadores más rigurosos consideran necesarias para que tengan significación clínica ( $> 0.75$ ; Fan et al., 2006). Sin embargo, un análisis de las coordenadas de las curvas comentadas no permitió encontrar ningún punto de corte que tuviera simultáneamente una sensibilidad y especificidad superiores al 70% (valor mínimo recomendado).

Lamentablemente, como en los tres únicos estudios publicados sobre el rendimiento diagnóstico de las escalas CC y D del MCMI-II (Millon, 2002; Piersma, 1991; Wetzler y Marlowe, 1993) no se informó del análisis de la curva ROC ni del área bajo la curva ROC, es imposible comparar los resultados sobre la curva ROC del presente

estudio con otros previos, lo cual señala la necesidad de replicarlos en futuras investigaciones y, a su vez, confirma la llamada de atención realizada por algunos investigadores sobre la necesidad de conocer una mayor amplitud de índices diagnósticos para establecer la validez diagnóstica de un instrumento (Hsu, 2002) como, por ejemplo, el área bajo la curva ROC el cual es considerado por muchos especialistas como el índice que mejor resume el rendimiento diagnóstico de un instrumento (Fan et al., 2006; Fischer et al., 2003; Streiner y Cairney, 2007; Swets, 1988).

**¿Son mejores los índices de validez diagnóstica de las escalas CC y D del MCMI-II que los de otros instrumentos ampliamente utilizados para evaluar la depresión como, por ejemplo, el BDI-II?**

Los resultados del presente estudio indican que los índices de validez diagnóstica de las escalas CC y D del MCMI-II para el diagnóstico del trastorno depresivo mayor o para el diagnóstico de cualquier trastorno depresivo no son mejores que los índices del BDI-II, sino más bien similares.

El área bajo la curva ROC de las puntuaciones del BDI-II para el diagnóstico del trastorno depresivo mayor fue, como las áreas de las puntuaciones directas de la escala D y de las puntuaciones tasa base de las escalas CC y D, estadísticamente significativa y superó el rango mínimo de las exactitudes aceptables ( $> 0.70$ ), pero fue la única que alcanzó exactitudes que otros investigadores más rigurosos consideran necesarias para que tengan significación clínica ( $> 0.75$ ; Fan et al., 2006). Cuando se excluyeron del análisis los pacientes con diagnósticos de otros trastornos depresivos en el cálculo del área de la curva ROC de las escalas con respecto al diagnóstico del trastorno depresivos mayor, el área de la curva ROC de las puntuaciones del BDI-II también fue

estadísticamente significativa y superó el límite más riguroso ( $> 0.75$ ), al igual que las puntuaciones tasa base y directas de la escala CC del MCMI-II. Por otro lado, del mismo modo que las escalas CC y D del MCMI-II, el área bajo la curva ROC del BDI-II para el diagnóstico de cualquier trastorno depresivo ni fue estadísticamente significativa ni siquiera alcanzó el valor mínimo que representa una exactitud aceptable ( $> 0.70$ ; Fischer et al., 2003; Streiner y Cairney, 2007; Swets, 1988).

De hecho, las pruebas estadísticas de comparación de pares de curvas ROC no revelaron diferencias estadísticamente significativas entre las áreas bajo las curvas ROC de las puntuaciones directas del BDI-II y las áreas bajo las curvas ROC de las puntuaciones directas o TB de las escalas CC y D del MCMI-II. Es más, los índices de validez diagnóstica de los diferentes puntos de corte evaluados para el BDI-II, incluyendo los índices de las puntuaciones del BDI-II que los autores del instrumento original han propuesto para diferenciar entre distintos niveles de gravedad de la depresión (Beck et al., 2011), se encontraban, en general, por debajo de los estándares aceptables para un instrumento diagnóstico, tal y como ocurría en el caso de las escalas CC y D del MCMI-II. Así, para todas las puntuaciones del BDI-II evaluadas los índices de concordancia diagnóstica (kappa) estaban por debajo del estándar de .40 y los índices de Youden no reflejaban un equilibrio adecuado entre sensibilidad y especificidad ( $< 40$ ), de manera que, en general, se encontraron índices de especificidad y de poder predictivo negativo adecuados para las puntuaciones del BDI-II que reflejan niveles moderados-graves de sintomatología depresiva, pero a costa de unos índices de sensibilidad y de poder predictivo positivo pobres; por otro lado, para las puntuaciones del BDI-II que reflejan niveles leves-moderados de sintomatología depresiva se alcanzaron índices de sensibilidad adecuados, pero a costa de índices de especificidad y de poder predictivo negativo pobres y también de índices de poder predictivo positivo pobres.

En resumen, los resultados del presente estudio indican que el rendimiento diagnóstico de las escalas CC y D del MCMI-II para el diagnóstico del trastorno depresivo mayor o para el diagnóstico de cualquier trastorno depresivo fue similar al del BDI-II y, para los tres instrumentos, dicho rendimiento fue cuestionable.

En el caso del BDI-II, estos datos de eficacia diagnóstica son inferiores a los encontrados en los estudios de la adaptación española del BDI-II (Beck et al., 2011; Sanz y García-Vera, 2013; Sanz et al., 2003). En estos estudios se analizó la eficacia diagnóstica del BDI-II en una muestra de estudiantes universitarios (Sanz et al., 2003) y en una muestra de pacientes con trastornos psicológicos (Sanz y García-Vera, 2013). En el estudio con estudiantes universitarios y tomando como referencia el diagnóstico alcanzado tras la aplicación de la entrevista diagnóstica estructurada computerizada Quick DIS-III-R (Bucholz, Marion, Shayka, Marcus y Robins, 1996), se encontró un área bajo la curva ROC de 0.91, lo que refleja una exactitud diagnóstica alta o bastante alta, y varias puntuaciones de corte con índices de concordancia diagnóstica (kappa) superiores a 0.50, con índices de Youden que reflejaban un equilibrio adecuado entre sensibilidad y especificidad ( $> 40$ ), con sensibilidades y especificidades simultáneas por encima del estándar de 70% y con índices de poder predictivo positivo superiores a 60% simultáneamente con índices de poder predictivo negativo superiores a 90%. Sin embargo, estos resultados tan positivos no deberían extrañar, ya que, lógicamente, la capacidad de discriminación diagnóstica de un instrumento depende de las características de la población en la que se está utilizado y, en este sentido, el diagnóstico diferencial en una muestra de pacientes con trastornos psicológicos entraña más dificultad que la detección de estudiantes universitarios con trastorno depresivo mayor en una muestra no clínica de estudiantes universitarios.

Más relevancia tienen, por tanto, los resultados obtenidos con la muestra de pacientes con trastornos psicológicos (Sanz y García-Vera, 2013) y, en este sentido, los resultados, aunque peores que los encontrados con estudiantes universitarios, también apoyaban empíricamente la validez diagnóstica de la adaptación española del BDI-II para discriminar entre pacientes con trastorno depresivo mayor y pacientes con otros trastornos mentales no depresivos (excluyendo, pues, a los pacientes con otros trastornos depresivos distintos al trastorno depresivo mayor). Así, el área bajo la curva ROC para el BDI-II fue de .83, lo que implica un rendimiento diagnóstico moderado a bueno (Fischer et al., 2003; Steiner y Cairney, 2007; Swets, 1988) o clínicamente significativo (Fan et al., 2006) y, además, se pudieron identificar varias puntuaciones de corte del BDI-II con índices de concordancia diagnóstica aceptables ( $\kappa > 0.40$ ), con índices de Youden que reflejaban un equilibrio adecuado entre sensibilidad y especificidad ( $> 40$ ), con sensibilidades y especificidades simultáneas por encima del estándar de 70%, y con índices de poder predictivo positivo superiores a 60% simultáneamente con índices de poder predictivo negativo superiores a 85%.

Curiosamente, los pacientes con trastornos psicológicos del estudio de Sanz y García-Vera (2013)<sup>1</sup> procedían de la misma Clínica Universitaria de Psicología de la UCM de la que fueron seleccionados los pacientes del presente estudio y los psicólogos que hicieron el diagnóstico de referencia también pertenecían al mismo colectivo que realizó el diagnóstico en este estudio: psicólogos residentes de la Clínica con formación de posgrado en psicología clínica y una experiencia clínica de al menos dos años. Quizás la diferencia más importante entre el estudio de Sanz y García-Vera (2013) y el presente estudio es que en aquél el criterio diagnóstico de referencia era el realizado por los psicólogos sobre la base, al menos, de la información de una entrevista diagnóstica

---

<sup>1</sup> La muestra de pacientes del estudio de Sanz y García-Vera (2013) se solapa parcialmente con la que participó en el estudio 2 de la presente tesis doctoral y que se presenta en el siguiente capítulo.

estructurada (el módulo de trastornos del estado de ánimo de la SCID-I VC; First, Spitzer, Gibbon y Williams, 1999) y de un listado de cotejo de síntomas depresivos (LED), mientras que en el presente estudio tan sólo hubo la seguridad de que todos los psicólogos utilizaron el listado de cotejo de síntomas depresivos (LED) para llegar al diagnóstico de referencia.

Esta diferencia metodológica permite plantear si los resultados de rendimiento diagnóstico tan cuestionables obtenidos en este estudio para las escalas CC y D del MCMI-II y por el BDI-II no podrían deberse a un problema en el criterio diagnóstico de referencia, más que a deficiencias diagnósticas de los propios instrumentos, esto es, del MCMI-II y del BDI-II. De hecho, en los estudios anteriores sobre el rendimiento diagnóstico de las escalas CC y D del MCMI-II (Millon, 2002; Piersma, 1991; Wetzler y Marlowe, 1993), los psicólogos o psiquiatras que realizaron el diagnóstico de referencia tampoco utilizaron ninguna entrevista diagnóstica estructurada, ni siquiera un listado de cotejo de síntomas depresivos.

Rogers (2003) revisa los datos de varios estudios que indican que en el proceso diagnóstico que realizan los clínicos (psicólogos o psiquiatras) a través de entrevistas no estructuradas o clínicas, existe un alto porcentaje de pacientes no diagnosticados con trastornos depresivos que sí son detectados y diagnosticados con dichos trastornos cuando, en cambio, los clínicos realizan los diagnósticos mediante entrevistas diagnósticas estructuradas. Por tanto, futuras investigaciones que traten de replicar los resultados del presente estudio deberían utilizar como criterio diagnóstico de referencia el diagnóstico realizado por un clínico sobre la base de la aplicación de una entrevista diagnóstica estructurada que, al fin y al cabo, es, hoy por hoy, lo más cercano a un “patrón oro” del diagnóstico en psicopatología, o al menos el procedimiento que asegura una mayor fiabilidad y validez en el diagnóstico (Segal, Coolidge, O’Riley y Heinz, 2006).

### **Limitaciones del presente estudio**

Además de la limitación señalada antes sobre la utilización como diagnósticos de referencia de diagnósticos no basados en entrevistas diagnósticas estructuradas, limitación de la que, tal y como se ha comentado, también adolecen los estudios previos sobre el rendimiento diagnóstico de las escalas CC y D del MCMI-II (Millon, 2002; Piersma, 1991; Wetzler y Marlowe, 1993), este estudio también presenta otras limitaciones que es importante señalar para matizar las conclusiones derivadas del mismo y para tratar de solventarlas, en la medida de lo posible, en futuras investigaciones, ya que muchas de estas limitaciones tienen que ver con su carácter retrospectivo.

En primer lugar, el diagnóstico de referencia se realizó no sólo sin aplicar una entrevista diagnóstica estructurada, sino también basado en un número reducido de sesiones de evaluación. En segundo lugar, los psicólogos que realizaron los diagnósticos podían conocer las puntuaciones del MCMI-II y del BDI-II antes de realizar los diagnósticos correspondientes, lo que podría influir en su toma de decisiones diagnósticas, aunque en este caso el sesgo sería a favor de la sobrevaloración de las cualidades diagnósticas de los instrumentos. De hecho, esta limitación es muy frecuente en los estudios sobre la validez diagnóstica de los instrumentos psicométricos y, en concreto, también estaba presente en dos de los tres estudios previos sobre el rendimiento diagnóstico de las escalas CC y D del MCMI-II (Piersma, 1991; Wetzler y Marlowe, 1993). Por tanto, en futuras investigaciones sobre la validez diagnóstica del MCMI-II ampliar el número de sesiones de evaluación que los clínicos realizan antes de llegar a un diagnóstico y que este diagnóstico se realice a partir, al menos, de una



entrevista estructurada diagnóstica y sin que el clínico conozca de antemano las puntuaciones de los pacientes en las escalas CC y D del MCMI-II.

### **6.5. Conclusiones**

Este es el primer estudio publicado sobre la validez diagnóstica de las escalas de D y CC de la adaptación española del MCMI-II, por lo que son necesarios nuevos estudios de replicación con diseños prospectivos que permitan superar las limitaciones ya señaladas y comprobar la generalizabilidad de los datos obtenidos sobre la validez diagnóstica de dichas escalas. En cualquier caso, y teniendo en cuenta sus limitaciones, los datos del presente estudio indican que las escalas D y CC de la adaptación española del MCMI-II presentan una validez diagnóstica cuestionable, especialmente en cuanto a su capacidad para detectar pacientes con trastorno depresivo mayor o pacientes con cualquier trastorno depresivo, aunque podría utilizarse en cierta medida para descartar la presencia de un trastorno depresivo. En consecuencia, se recomienda, en tanto en cuanto no se realicen investigaciones más exhaustivas sobre el rendimiento diagnóstico de la adaptación española del MCMI-II, no con, sino que tal diagnóstico debería estar basado en otras fuentes de información, obtenida a ser posible a través de la aplicación de una entrevista diagnóstica estructurada, sobre todo si el objetivo es realizar un diagnóstico psicopatológico.

**Capítulo 7. SEGUNDO ESTUDIO:  
VALIDEZ DIAGNÓSTICA DE LAS  
ESCALAS DE DEPRESIÓN  
MAYOR Y DISTIMIA DEL MCMI-III**



### 7.1. Introducción

Como se ha comentado en el capítulo 6 de esta tesis doctoral, el MCMI es uno de los cuestionarios más utilizados en psicología en Europa (Evers et al., 2012) y especialmente en España (Muñiz y Fernández-Hermida, 2010). Sin embargo, aunque uno de los principales objetivos de este inventario es la realización de diagnósticos (Millon, Davis y Millon, 2007), no existe suficiente información sobre sus propiedades psicométricas como instrumento diagnóstico, incluyendo las propiedades para diagnosticar trastornos depresivos de sus escalas de Depresión Mayor (CC) y Distimia (D), situación que también afecta a la última versión del MCMI hasta la fecha: el MCMI-III.

Efectivamente, a pesar de los casi 20 años transcurridos desde que el MCMI-III fue editado por primera vez (Millon, 1994), tan sólo se han podido localizar cinco estudios publicados que ofrecen datos empíricos sobre la eficacia de las escalas de Depresión Mayor y Distimia del MCMI-III para diagnosticar trastornos depresivos, cuatro de ellos sobre la versión original estadounidense, incluyendo los dos realizados por los propios autores del instrumento, y uno sobre la adaptación belga en población flamencoparlante (Gibeau y Choca, 2005; Millon, 1994; Millon et al., 1997; Rossie y Sloore, 2005; Saulsman, 2011) (véanse las Tablas 7.1 y 7.2).

Existe, además, una gran controversia en torno a la eficacia diagnóstica de la versión original del MCMI-III (Hsu, 2002). Los datos proporcionados por los autores en el primer manual del MCMI-III (Millon, 1994; véase Hsu, 2002; Rezlaff, 1996) eran significativamente inferiores a los publicados en la segunda edición del manual (Millon et al., 1997). En el caso de esta segunda edición del manual (véase la Tabla 7.1), los datos proporcionados para la puntuación de corte de tasa base 85 indicaban, con respecto al diagnóstico de sus respectivas etiquetas diagnósticas, una sensibilidad por encima del 70% para la escala de Depresión Mayor, superior al mínimo aceptable en un instrumento

diagnóstico (Lenderink et al., 2012; Sheehan et al., 1998), una sensibilidad de 65% para la escala de Distimia y unos índices kappa superiores a 0.6 para ambas escalas, lo que se considera buenos índices kappa (Altman, 1991). Sin embargo, algunos autores han señalado que los procedimientos que se llevaron a cabo para la realización de esos estudios tenían serios problemas metodológicos (Grove y Vrieze, 2009).

**Tabla 7.1.** Índices de validez diagnóstica de las escalas de Depresión Mayor y Distimia de la versión original del MCMI-III en los estudios de validez realizados para las dos primeras ediciones de su manual

Índices	Millon (1994) <sup>a</sup> (preliminares)		Millon (1994) <sup>b</sup>				Millon et al. (1997) <sup>c</sup>	
	Depresión	Distimia	Depresión		Distimia		Depresión	Distimia
	TB85	TB85	TB75	TB85	TB75	TB85	TB85	TB85
Prevalencia	21.5	36.2	20	20	38	38	11	23
Sensibilidad	56.6	36.7	85	65	85	55	84	65
Especificidad	73	58.8						
VPP	36	34	51	72	61	88	66	81
VPN	86	62					98	90
Eficacia								
Kappa	0.24	-0.04					0.70	0.65
Youden <sup>d</sup>	29.4	-4.5						

*Nota.* VPP: valor predictivo positivo. VPN: valor predictivo negativo. Eficacia: índice de eficacia global o porcentaje de pacientes correctamente clasificados (también denominado precisión, índice de validez o de eficiencia, o poder o valor diagnóstico total).

<sup>a</sup>Estadísticos de eficacia para el diagnóstico del trastorno depresivo mayor en el caso de la escala de Depresión Mayor y para el diagnóstico del trastorno distímico en el caso de la escala de Distimia, calculados por Rezlaff (1996) y Hsu (2002) sobre los datos preliminares de Millon (1994).

<sup>b</sup>Estadísticos de eficacia para el diagnóstico del trastorno depresivo mayor en el caso de la escala de Depresión Mayor y para el diagnóstico del trastorno distímico en el caso de la escala de Distimia; datos definitivos de la primera edición del manual del MCMI-III (Millon et al., 1997).

<sup>c</sup>Estadísticos de eficacia para el diagnóstico del trastorno depresivo mayor en el caso de la escala de Depresión Mayor y para el diagnóstico del trastorno distímico en el caso de la escala de Distimia, calculados por Hsu (2002) sobre los datos de la segunda edición del manual del MCMI-III (Millon et al., 1997).

<sup>d</sup>Este índice no aparece recogido en la publicación correspondiente, sino que ha sido calculado para este estudio a partir de los datos de la tabla.

Por otro lado, se han publicado dos estudios independientes sobre la eficacia diagnóstica de las escalas de Depresión Mayor y Distimia de la versión original del MCMI-III (Gibeau y Choca, 2005; Saulsman, 2011). En ambos estudios (véase la Tabla 7.2), se encontraron sensibilidades menores a 70% (excepto para la puntuación TB 75 de la escala de Distimia en el estudio de Gibeau y Choca, 2005), inferiores a lo recomendado como aceptable para un instrumento diagnóstico (Lenderink et al., 2012; Sheehan et al., 1998), aunque, en general, sí se encontraron especificidades superiores a 70% (excepto para la puntuación TB 75 de la escala de Distimia en ambos estudios). Por tanto, no es de extrañar que los índices de Youden calculados sobre esos datos mostraran un desequilibrio entre sensibilidad y especificidad con valores que, en general, eran inferiores a 40 (salvo para la puntuación TB 75 de la escala de Depresión Mayor en el estudio de Saulsman, 2011; véase la Tabla 7.2); además, en ambos estudios no se publicaron otros estadísticos de su eficacia diagnóstica como, por ejemplo, el índice kappa, que hubiesen permitido una valoración más sólida de la eficacia diagnóstica de las escalas de Depresión Mayor y Distimia del MCMI-III. En cualquier caso, estos estudios ofrecen más información sobre el rendimiento diagnóstico de las escalas de la versión original inglesa del MCMI-III y permiten que los usuarios de dicha versión puedan valorar ese rendimiento de una manera más independiente y más allá de los estudios de los propios autores del instrumento.

En la adaptación belga para flamencoparlantes del MCMI-III (Rossie y Sloore, 2005), los autores seleccionaron puntos de corte que tuvieran una sensibilidad mayor del 70% para la asignación de la puntuación TB 75 de las diferentes escalas, aunque la especificidad de las escalas de Depresión Mayor y Distimia no superó el 60% ni para las puntuaciones TB 75 ni para las puntuaciones TB 85 (véase la Tabla 7.2). En este caso, sí se calcularon otros estadísticos, como el índice kappa, inferiores en ambas escalas a 0.30 para ambos tipos de puntuaciones (índice de kappa débil; Altman, 1991) y, de

nuevo, los índices de Youden calculados sobre esos datos mostraron un desequilibrio entre sensibilidad y especificidad con valores inferiores a 40.

**Tabla 7.2.** Índices de validez diagnóstica de las escalas de Depresión Mayor y Distimia del MCMI-III en estudios independientes con muestras heterogéneas de pacientes con trastornos psicológicos

Índices	Gibeau y Choca (2005) <sup>a</sup>				Saulsman (2011) <sup>b</sup>				Rossi y Sloore (2005) <sup>c</sup>			
	Depresión		Distimia		Depresión		Distimia		Depresión		Distimia	
	TB75	TB85	TB75	TB85	TB75	TB85	TB75	TB85	TB75	TB85	TB75	TB85
Prev.	42	42	42	42	58	58	58	58	47	29	42	25
Sens.	45	26	73	32	55	35	66	21	73	73	70	66
Espec.	79	90	58	86	87	91	64	95	51	55	47	60
VPP	60	67	56	62	85	84	72	85	57	10	50	36
VPN	67	63	75	63	58	50	58	46	68	83	69	83
Eficacia	65	63	65	63	—	—	—	—	61	60	57	61
Kappa	—	—	—	—	—	—	—	—	0.24	0.26	0.16	0.2
Youden <sup>d</sup>	23	16	31	18	42	26	30	16	24	28	17	26

Nota. Prev.: prevalencia. Sens.: sensibilidad. Espec.: especificidad. VPP: valor predictivo positivo. VPN: valor predictivo negativo. Eficacia: índice de eficacia global o porcentaje de pacientes correctamente clasificados (también denominado precisión, índice de validez o de eficiencia, o poder o valor diagnóstico total).

<sup>a</sup>Eficacia para el diagnóstico del conjunto de trastornos depresivos, por eso la prevalencia que se recoge se refiere a dicho conjunto (trastorno depresivo mayor, distimia, trastorno depresivo no especificado, etc.).

<sup>b</sup>Eficacia para el diagnóstico de los trastornos depresivos sin comorbilidad frente a un grupo control de pacientes con diagnóstico de trastorno de ansiedad sin comorbilidad, por eso la prevalencia que se recoge se refiere a la suma de las frecuencias de los trastornos depresivos en esa muestra de pacientes sin comorbilidad (45.4% de trastorno depresivo mayor + 12.6% de distimia).

<sup>c</sup>Eficacia para el diagnóstico del trastorno depresivo mayor en el caso de la escala de Depresión Mayor y para el diagnóstico del trastorno distímico en el caso de la escala de Distimia.

<sup>d</sup>Este índice no aparece recogido en la publicación correspondiente, sino que ha sido calculado para este estudio a partir de los datos de la tabla.

No se ha publicado ningún estudio sobre la eficacia diagnóstica de las escalas de Depresión Mayor (CC) y Distimia (D) del MCMI-III en su adaptación española (Millon, Davis y Millon, 2007, adaptación española de V. Cardenal y M. P. Sánchez). Aunque el procedimiento para el cálculo de las escalas Tasa Base (TB) en la adaptación española

del inventario fue similar al empleado en la elaboración de su versión original, no se puede asumir automáticamente que los puntos de corte TB 75 y TB 85 tengan adecuados índices de eficacia diagnóstica, máxime cuando los resultados de los estudios previos realizados con la versión original del MCMI-III o con otras adaptaciones no son del todo satisfactorios respecto a la validez diagnóstica de sus escalas CC y D (Gibeau y Choca, 2005; Hsu, 2002; Rezlaff, 1996; Rossi y Sloore, 2005; Saulsman, 2011) y plantean dudas sobre si la utilización de las puntuaciones tasa base es el mejor procedimiento para establecer los puntos de corte diagnósticos para el MCMI-III (Rossi y Sloore, 2005).

Por tanto, para estudiar la eficacia diagnóstica de las escalas de Depresión Mayor y Distimia del MCMI-III en su adaptación española (Millon et al., 2007, adaptación española de V. Cardenal y M. P. Sánchez) se llevó a cabo este segundo estudio empírico, el cuál fue diseñado como un estudio prospectivo que permitiera controlar los principales sesgos que pueden presentarse en las investigaciones sobre el rendimiento diagnóstico de los instrumentos de medida, de manera que el evaluador no tuviera acceso a las puntuaciones en los instrumentos a valorar hasta que no hubiese alcanzado el diagnóstico del paciente.

Por otro lado, con el objetivo de aumentar la validez del diagnóstico clínico utilizado como criterio de referencia, se aplicó a los pacientes una entrevista diagnóstica estructurada (la SCID-I) y los clínicos realizaron al menos tres entrevistas clínicas de evaluación de una hora de duración con el paciente antes de llegar a un diagnóstico, diagnóstico este último que reflejaron en una ficha que rellenaban al terminar el periodo de evaluación. De esta manera, se contó con dos criterios de referencia: el diagnóstico realizado por el clínico tras administrar la SCID-I y el diagnóstico realizado por el clínico al finalizar el período de evaluación (que implicaba al menos tres entrevistas clínicas) y que anotaba en su ficha diagnóstica o clínica.



Estos dos diagnósticos sirvieron como criterio de referencia para evaluar la eficacia diagnóstica de las puntuaciones TB 75 y TB 85 de las escalas CC y D del MCMI-III respecto a su capacidad para detectar, en primer lugar, pacientes con trastorno depresivo mayor y, en segundo lugar, pacientes con todo tipo de trastornos depresivos, incluyendo en este último grupo no sólo los dos tipos principales considerados por el DSM-IV (trastorno depresivo mayor y trastorno distímico), sino también todas las categorías que el DSM-IV considera residuales y relacionadas con la depresión (trastorno adaptativo con estado de ánimo depresivo, duelo y trastorno depresivo no especificado). Además se estudió la eficacia diagnóstica de otros puntos de cortes alternativos a las puntuaciones TB 75 y TB 85 de las escalas de los trastornos depresivos del MCMI-III.

Por otro lado, como un objetivo secundario y tal como se hizo en el primer estudio de esta tesis (véase el capítulo 6), el presente estudio también pretendía comparar la eficacia diagnóstica del MCMI-III frente al Inventario de Depresión de Beck-II (BDI-II).

Finalmente, siguiendo la recomendación de Ganellen (1996) ya comentada en el capítulo 6 de esta tesis, se estudió la eficacia diagnóstica de las escalas de Depresión Mayor y Distimia del MCMI-III para diagnosticar el trastorno depresivo mayor no solamente frente al resto de trastornos mentales, sino también frente a los trastornos mentales no depresivos, excluyendo, por tanto, del grupo de contraste de pacientes con otros trastornos mentales, a los pacientes con trastorno distímico, trastorno adaptativo con estado de ánimo depresivo, duelo o trastorno depresivo no especificado, así como a los pacientes con un diagnóstico secundario de trastorno depresivo.

## 7.2. Método

### 7.2.1. Participantes

En el presente estudio participaron 203 pacientes con trastornos psicopatológicos atendidos en la Clínica Universitaria de Psicología de la UCM, que tenían más de 18 años de edad, a los que se les había aplicado el MCMI-III y cuyas puntuaciones en sus índices de validez no cumplían las condiciones de invalidación de la prueba establecidas en su manual (Millon, Davis y Millon, 2007, adaptación española de V. Cardenal y M. P. Sánchez, p. 123), razón por la que tres pacientes quedaron excluidos del estudio. La edad media de la muestra fue 31 años ( $DT = 11.6$ ). En las Tablas 7.3, 7.4, 7.5, 7.7 y 7.7 se describen las principales características sociodemográficas (sexo, estado civil, nivel de estudios y situación laboral) y clínicas (demanda asistencial) de la muestra.

**Tabla 7.3.** Sexo en la muestra del estudio 2

Sexo	Frecuencia	Porcentaje
Varones	61	30%
Mujeres	140	69%
Total	201	99%
Perdidos	2	1%

**Tabla 7.4.** Estado civil en la muestra del estudio 2

Estado Civil	Frecuencia	Porcentaje
Soltero	132	65
Casado/ Convivencia	48	23.6
Viudo	2	1
Separado/divorciado	17	8.4
Total	199	98.0
Sin datos	4	2.0

**Tabla 7.5.** Nivel de estudios en la muestra del estudio 2

<b>Nivel de Estudios</b>	<b>Frecuencia</b>	<b>Porcentaje</b>
Estudios primarios	5	2.5
EGB	18	8.9
Bachillerato	64	31.5
Universitarios	112	55.2
Otros	1	0.5
Total	200	98.5
Sin datos	3	1.5

**Tabla 7.6.** Profesión/situación laboral en la muestra del estudio 2

<b>Profesión/situación laboral</b>	<b>Frecuencia</b>	<b>Porcentaje</b>
Trabaja	100	49.3
Jubilado o pensionista	2	1.0
Parado/a	7	3.4
Estudiante	83	40.9
Sus labores	4	2.0
Otra situación	4	2.0
Total	200	98.5
Sin datos	3	1.5

**Tabla 7.7.** Demanda asistencial en la muestra del estudio 2

<b>Demanda Asistencial</b>	<b>Frecuencia</b>	<b>Porcentaje</b>
Tristeza, bajo estado de ánimo	27	13.3
Ansiedad	58	28.6
Ansiedad + tristeza	23	11.3
Baja autoestima	2	1.0
Problemas de relación	12	5.9
Depresión + otros	13	6.4
Tristeza + baja autoestima	8	3.9
Ansiedad + otros	14	6.9

<b>Demanda Asistencial</b>	<b>Frecuencia</b>	<b>Porcentaje</b>
Otros	38	18.7
Ruptura de pareja/duelo	6	3.0
Total	201	99.0
Sin datos	2	1.0

Como se ha comentado en el capítulo 6 de esta tesis, la Clínica Universitaria de Psicología de la UCM es un centro que presta atención psicológica ambulatoria a estudiantes y personal de la universidad, pero también a personas ajenas a la misma (García-Vera y Sanz, 2009), y, quizás por esa razón, el perfil de pacientes y problemas clínicos que atiende es comparable en muchos aspectos al que se puede encontrar en centros privados españoles de psicología clínica (Estupiñá et al., 2012). De hecho, en lo que concierne a la presencia de sintomatología y trastornos depresivos, la muestra de pacientes del presente estudio se asemejaba más a la población española de pacientes ambulatorios con trastornos mentales que la muestra del estudio primero (véase el capítulo 6). Así, la muestra de pacientes del presente estudio obtuvo una puntuación media en el BDI-II de 21.74 ( $DT = 11.07$ ), similar a la que se ha encontrado en otras investigaciones con muestras de pacientes psicopatológicos, ya que en una revisión reciente de las características psicométricas del BDI-II en muestras de pacientes psicopatológicos de diversos países, se calculó una puntuación media de 22.8 ( $DT = 11.9$ ) promediando de forma ponderada los resultados de 10 estudios que totalizaban 2 978 pacientes (Sanz y García-Vera, 2009). Lo que es más importante, las medianas de las puntuaciones directas en la escala de Depresión Mayor del MCMI-III de la presente muestra eran muy similares a las TB 60 de la muestra de tipificación de la adaptación española del MCMI-III, las cuales, por definición, representan la mediana de esa muestra psicopatológica de tipificación. Así, las medianas de la escala de Depresión Mayor del MCMI-III en los pacientes varones y mujeres de esta investigación fueron,

respectivamente, 7 y 8, mientras que las TB 60 de la muestra de tipificación española de pacientes varones y mujeres fueron 5 y 8 (Millon et al., 2007, adaptación española de V. Cardenal y M. P. Sánchez).

### 7.2.2. Instrumentos

En la presente investigación se utilizaron los siguientes instrumentos:

**1. MCMI-III** (Millon, 1994) **en su adaptación española** (Millon, Davis y Millon, 2007, adaptación española de V. Cardenal y M. P. Sánchez). El MCM-III es un instrumento de autoinforme que consta de 175 ítems de respuesta verdadero-falso y que permite obtener medidas en 4 escalas de control, 11 escalas básicas de patrones clínicos de personalidad, 3 escalas de patología graves de personalidad, 7 síndromes clínicos de gravedad moderada y 3 síndromes de gravedad más extrema (véase la descripción de las escalas en el Anexo 7.1). De todas las escalas del MCMI-III, en el presente estudio se utilizaron las escalas CC y D del MCMI-III y dos escalas que señalan hasta qué punto el perfil del MCMI-III se considera válido o inválido. Estas dos escalas son las siguientes (Millon et al., 2007):

- **Sinceridad** (escala X): Indica si el paciente se inclina a ser reservado y reticente, o por el contrario es franco y revelador. Se obtuvo de la suma de las puntuaciones en las escalas clínicas e invalida el cuestionario en el caso de que existan puntuaciones por debajo de 34 o por encima de 178.
- **Validez** (escala V): incluye 3 ítems con posibles respuestas inverosímiles, que de responderse verdadero, indican que el cuestionario debe interpretarse con cautela.

Siguiendo los criterios establecidos en el manual del MCMI-III (Millon et al., 2007), en el presente estudio se incluyeron aquellos pacientes que puntuaron  $< 2$  en la escala V y  $> 34$  y  $< 179$  en la escala X. Como se ha comentado, en este estudio 3

pacientes no cumplieron estos criterios y por tanto quedaron excluidos. Una descripción más detallada del MCMI-III, de sus escalas CC y D y de sus propiedades psicométricas se puede consultar en el capítulo 4 de esta tesis doctoral.

## **2. Entrevista Clínica Estructurada para los Trastornos del Eje I del DSM-IV, Versión**

**Clínica (SCID-I;** First, Spitzer, Gibbon y Williams, 1999). Esta entrevista semiestructurada se desarrolló para el diagnóstico de los trastornos mentales del Eje I del DSM-IV. Es probablemente la entrevista más utilizada en investigación en EE. UU. y, de hecho, “la entrevista SCID en varios de sus formatos, ha sido utilizada en más de 1000 estudios publicados” (Summerfeldt, Antony y Klossterman, 2011). Se utilizó en este caso el módulo específico de trastornos afectivos que permite la valoración del episodio depresivo mayor, episodio maniaco, episodio hipomaniaco, distimia y trastornos del estado de ánimo debido al consumo de sustancias o debido a enfermedades médicas. Es la entrevista con más soporte empírico para el diagnóstico de los trastornos del estado de ánimo (Joiner, Walker, Pettit, Pérez y Cukrowicz, 2005).

## **3. Inventario de Depresión de Beck, segunda edición (BDI-II; Beck et al., 1996) en su**

**adaptación española** (Sanz et al., 2005; Sanz y García-Vera, 2013; Sanz, Navarro y Vázquez, 2003; Sanz, Perdigón y Vázquez, 2003) que recientemente ha sido publicada por *Pearson Talent and Assessment España* (Beck, Steer y Brown, 2011a, adaptación de J. Sanz y C. Vázquez). Una descripción más detallada del BDI-II se puede consultar en el capítulo 3 de la presente tesis así como en el apartado homólogo del capítulo 6.

### 7.2.3. Procedimiento

La aplicación de los instrumentos se llevó a cabo desde mayo de 2008 a mayo de 2010. Tras ponerse en contacto con la Clínica Universitaria de Psicología de la UCM, cada paciente fue evaluado durante 4 sesiones de una hora de duración (véase la Tabla 7.8) Los psicólogos aplicaron los cuestionarios MCMI-III y BDI-II durante las 2 primeras sesiones de evaluación, pero, hasta completar el procedimiento de la investigación con cada paciente, no pudieron consultar ni las puntuaciones resultantes de su corrección ni las respuestas ítem a ítem que habían dado los pacientes. Los psicólogos administraron el módulo de trastornos del estado de ánimo de la SCID-I durante la segunda o la tercera sesión con cada paciente, con el objetivo de tener una información más fiable y válida sobre un posible diagnóstico depresivo del paciente. Durante las cuatro sesiones de evaluación los psicólogos aplicaron al paciente entrevistas semiestructuradas o libres, cuestionarios y otros instrumentos, que variaban de un paciente a otro en función de sus problemas, diagnósticos provisionales y preferencias de los clínicos.

Con la información obtenida a lo largo de las sesiones de evaluación, el psicólogo debía rellenar tras la 3ª o 4ª entrevista con el paciente una ficha de información sobre el mismo (véase el Anexo 6) en la que se le preguntaba por el posible diagnóstico del paciente y el grado de certeza sobre dicho diagnóstico, así como otros datos relevantes para el estudio. Además, los psicólogos debían valorar la gravedad de la sintomatología depresiva de los pacientes en las siguientes categorías: sin depresión, depresión leve, moderada o grave.

La recogida de los datos se realizó de modo centralizado. Una persona del equipo de la Clínica Universitaria de Psicología se encargó de corregir los MCMI-III y los BDI-II aplicados por los evaluadores, así como de codificar sus puntuaciones en una base de datos creada para tal efecto. Los psicólogos sólo pudieron conocer los resultados del

MCMI-III y del BDI-II una vez habían realizado la SCID-I y habían cumplimentado la ficha diagnóstica, de forma que no hubiera la posibilidad de sesgos o contaminación cruzada entre las puntuaciones de estos cuestionarios y los diagnósticos realizados por los psicólogos.

**Tabla 7.8.** Procedimiento de evaluación diagnóstica del estudio segundo

Instrumentos de medida	1ª sesión	2ª sesión	3ª sesión	4ª sesión
Aplicación del MCMI-III y del BDI-II				
Aplicación de la SCID-I				
Cumplimentación de la ficha diagnóstica				

#### 7.2.4. Evaluadores

Para cada paciente, un único evaluador, su terapeuta, era el encargado de aplicar todo el procedimiento de evaluación. Todos los terapeutas fueron psicólogos becarios residentes de la Clínica Universitaria de Psicología de la Universidad Complutense de Madrid, con formación de postgrado en psicología clínica y de la salud, y como mínimo con 2 años de experiencia clínica. Todos ellos, además, fueron entrenados en la aplicación de los instrumentos de evaluación descritos y en los pasos a seguir para la correcta aplicación de los mismos.

Todos los evaluadores recibieron un entrenamiento específico en la utilización de la entrevista estructurada SCID-I, como marcan las directrices de la misma entrevista (First, Spitzer, Gibbon, Williams y Smith, 1999), teniendo en cuenta que la fiabilidad y validez de esta entrevista depende del criterio clínico de los psicólogos y de su destreza a



la hora de aplicarla. Especialmente importante en la SCI-I es la profundización mediante preguntas en aquellas cuestiones que los clínicos crean relevantes, tomando las decisiones diagnósticas en función de toda la información disponible por el clínico para así aumentar su validez (First et al., 1999). Todos los psicólogos recibieron información escrita acerca de las características de la entrevista SCID-IV, solicitándoles la lectura de las Normas de Utilización y las Recomendaciones de la Guía del Usuario (First et al., 1999). Así mismo, se plantearon una serie de sesiones de *role-playing* con el objetivo de practicar algunos casos concretos y aspectos de la entrevista de difícil manejo.

#### **7.2.5. Diseño y análisis de datos**

El diseño de esta investigación responde al de un estudio instrumental de análisis de las propiedades psicométricas de un instrumento y, en concreto, al de un estudio de validación diagnóstica prospectiva, ya que se llevó a cabo en pacientes todavía sin diagnosticar a los que, tras la aplicación de los instrumentos a validar, se siguió hasta alcanzar un diagnóstico definitivo por otros medios, permitiendo evaluar adecuadamente la capacidad predictiva-diagnóstica del instrumento. Como comenta Carnero-Pardo (2005), el carácter prospectivo y el hecho de aplicar el instrumento antes de realizar el diagnóstico, permiten el control de los principales sesgos que afectan a los estudios de validación, pudiendo asegurar tanto la evaluación independiente y ciega del instrumento a validar y del criterio diagnóstico con el que se compara, como el que todos pacientes, independientemente de sus resultados, se sometan al proceso diagnóstico completo.

Los estadísticos utilizados para el análisis de la validez diagnóstica de los instrumentos fueron: sensibilidad, especificidad, índice de Youden, poder predictivo positivo, poder predictivo negativo, índice kappa, eficacia global y área bajo la curva ROC. Estos estadísticos están descritos en el capítulo 6 de esta tesis doctoral.

Todos estos estadísticos de eficacia o rendimiento diagnóstico fueron calculados para realizar las siguientes discriminaciones a partir de los puntos de corte propuestos por los manuales del MCMI-III y del BDI-II (p. ej., puntuaciones TB 75 y 85 en el caso del MCMI-III) u otros sugeridos por los propios estadísticos de rendimiento diagnóstico:

- 1) Entre pacientes con un diagnóstico de trastorno depresivo mayor y pacientes con otro tipo de diagnóstico según el DSM-IV.
- 2) Entre pacientes con cualquier diagnóstico del conjunto de trastornos depresivos (trastorno depresivo mayor, trastorno distímico, trastorno depresivo no especificado, trastorno adaptativo con estado de ánimo depresivo, trastorno adaptativo mixto y duelo) y pacientes con un diagnóstico no depresivo.
- 3) Entre pacientes con un diagnóstico de trastorno depresivo mayor y pacientes con un diagnóstico no depresivo (excluyendo, por tanto, entre estos últimos pacientes, aquéllos con un diagnóstico de otros trastornos depresivos: trastorno distímico, trastorno depresivo no especificado, trastorno adaptativo con estado de ánimo depresivo, trastorno adaptativo mixto y duelo, o con un diagnóstico secundario de trastorno depresivo).

Como se puede observar, durante el proceso de evaluación se recogían dos diagnósticos diferenciados. En primer lugar el diagnóstico asignado por el terapeuta tras el periodo completo de evaluación, sin conocer el resultado de los cuestionarios, pero sí habiendo aplicado el módulo de trastornos del estado de ánimo de la entrevista SCID-I (en adelante *diagnóstico de la ficha clínica*) y en segundo lugar, el diagnóstico realizado por los clínicos en la propia aplicación del módulo de trastornos del estado de ánimo de la entrevista SCID-I (en adelante *diagnóstico de la SCID-I*). En principio, el diagnóstico que

se pensaba utilizar como criterio de referencia para evaluar la validez diagnóstica de las escalas CC y D del MCMI-III y del BDI-II era el diagnóstico de la ficha clínica, pero al detectarse discrepancias entre ese criterio diagnóstico y el diagnóstico de la SCID-I, también se analizó el rendimiento diagnóstico de las escalas de CC y D del MCMI-III y del BDI-II utilizando como criterio el diagnóstico de trastorno depresivo mayor proporcionado por la SCID-I.

Las curvas ROC y los diferentes índices de rendimiento diagnóstico fueron calculados mediante el paquete estadístico IBM SPSS, versión 19. Para comparar la validez diagnóstica de las diferentes medidas, escalas e instrumentos (puntuaciones directas y TB de las escalas CC y D del MCMI-III y puntuaciones directas del BDI-II), se compararon las curvas ROC correspondientes mediante el programa estadístico EPIDAT, versión 3.1 (Xunta de Galicia y Organización Panamericana de la Salud, Santiago de Compostela, España; <http://dxsp.sergas.es>) y se compararon dos a dos las áreas de las curvas ROC mediante el estadístico z del programa estadístico MedCalc para Windows, versión 12.4.0 (MedCalc Software, Mariakerke, Bélgica; <http://www.medcalc.org>).

### **7.3. Resultados**

#### **Diagnósticos según el DSM-IV en la muestra de pacientes del estudio**

Según el diagnóstico que los clínicos asignaron a los pacientes en la ficha diagnóstica, la prevalencia del trastorno depresivo mayor en este estudio fue del 20.7%, mientras que la prevalencia de la distimia fue del 4.4% (véase la Tabla 7.9). En otros estudios realizados en centros de salud mental en España se han encontrado porcentajes muy similares de diagnósticos de trastornos depresivos, con una prevalencia del 22% (de Francisco y Goenaga, 2002) o del 23.9% (Valero y Ruiz, 2003). En la muestra de este estudio hubo una prevalencia del trastorno adaptativo con estado de ánimo depresivo del

3.4% y del 10.8% para el trastorno adaptativo mixto. Por tanto, la prevalencia del conjunto de los trastornos depresivos fue del 39.3%.

**Tabla 7.9.** Diagnósticos según el DSM-IV en la muestra del estudio segundo

Diagnóstico	Frecuencia	Porcentaje
Trastorno depresivo mayor	42	20.7
Trastorno distímico	9	4.4
Trastorno psicótico	4	2.0
Trastorno de angustia con agorafobia	8	3.9
Fobia social	10	4.9
Fobia específica	6	3.0
Trastorno de angustia sin agorafobia	8	3.9
Trastorno de ansiedad generalizada	9	4.4
Trastorno de ansiedad no especificado	11	5.4
Trastorno obsesivo-compulsivo	12	5.9
Trastorno adaptativo con estado de ánimo depresivo	7	3.4
Trastorno adaptativo mixto con ansiedad y estado de ánimo depresivo	22	10.8
Trastorno adaptativo con ansiedad	2	1
Anorexia	1	0.5
Bulimia	2	1
Otros trastornos de alimentación	4	2
Insomnio primario	2	1
Otros diagnósticos	16	7.9
Sin diagnóstico en el Eje I	24	11.8
Agorafobia sin historia de trastorno angustia	2	1
Consumo de sustancias	2	1
Total	203	100

Según el diagnóstico proporcionado por el módulo de los trastornos del estado de ánimo de la entrevista SCID-I (véase la Tabla 7.10), la prevalencia del trastorno depresivo

mayor fue de 37.5% y del trastorno distímico de 4.9%. Algunos autores han afirmado que existe un alto porcentaje de personas con un trastorno depresivo que no son detectadas durante las entrevistas clínicas no estructuradas, pero a las que sí se les asigna este diagnóstico cuando se llevan a cabo entrevistas diagnósticas semiestructuradas o estructuradas (Rogers, 2003), en este caso la SCID-I. Efectivamente, en este estudio, el porcentaje de personas diagnosticadas con un trastorno depresivo fue mayor teniendo en cuenta los resultados de la entrevista diagnóstica SCID-I que teniendo en cuenta el diagnóstico recogido por el clínico en la ficha diagnóstica del paciente.

**Tabla 7.10.** Diagnósticos de trastornos del estado de ánimo según la entrevista diagnóstica estructurada SCID-I en la muestra del estudio segundo

Diagnóstico	Frecuencia	Porcentaje
Trastorno depresivo mayor	69	37.5
Trastorno distímico	9	4.9
Trastorno depresivo subumbral	26	14.1
Duelo	1	0.5
Trastornos bipolares (sin episodio actual depresivo)	1	0.5
Sin sintomatología depresiva ni bipolar	78	42.4
Total	184	100

Con el fin de analizar las discrepancias entre el diagnóstico de la SCID-I y el diagnóstico de la ficha clínica se realizó un análisis de frecuencias. Se encontró que el porcentaje de acuerdo respecto al diagnóstico de trastorno depresivo mayor entre la SCID-I y la ficha clínica fue del 75.7%. Posteriormente, se realizó un análisis de frecuencias de las discrepancias encontradas (24.3% de los pacientes) para conocer su posible origen. En la Tabla 7.11 se puede observar que más del 80.4% de estas discrepancias consistían en pacientes que según la SCID-I sufrían un trastorno depresivo mayor, pero que tenían un diagnóstico de otro trastorno en su ficha diagnóstica. Los

grupos de trastornos más frecuentes en estas discrepancias fueron los trastornos de ansiedad (30.4%), el trastorno adaptativo mixto (23.9) y el trastorno distímico (10.9).

**Tabla 7.11.** Análisis de frecuencias de las discrepancias entre el diagnóstico SCID-I para el trastorno depresivo mayor y el diagnóstico de la ficha clínica

Diagnóstico SCID-I	Diagnóstico clínico	Frecuencia	Porcentaje
Trastorno depresivo mayor	Trastorno distímico	5	10.9
Trastorno depresivo mayor	Trastorno de ansiedad	14	30.4
Trastorno depresivo mayor	Trastorno adaptativo mixto	11	23.9
Trastorno depresivo mayor	Otros trastornos	7	15.2
Sin trastorno depresivo mayor	Trastorno depresivo mayor	9	19.6
Total de discrepancias		46	100

Con el fin de profundizar más en el análisis de estas discrepancias se revisaron de manera aleatoria los historiales clínicos de 13 de estos pacientes los cuales no tenían un trastorno depresivo mayor recogido en su ficha diagnóstica. Se analizaron los diagnósticos asignados en los informes clínicos (que se realizaron tras la finalización del periodo de evaluación) y en el que también figuraban las puntuaciones de todos los instrumentos aplicados. El 38.5% de los pacientes analizados (5 de estos 13 pacientes) tenían en dicho informe un diagnóstico de trastorno depresivo mayor, coincidiendo por tanto con el diagnóstico recogido en la aplicación de la entrevista SCID-I, pero no con el diagnóstico que figuraba en la ficha. Quizá los terapeutas, tras la redacción del informe clínico completo, al haber tenido en cuenta más información clínica, el resultados de los

cuestionarios y de la entrevista SCID-I, podrían haber realizado un diagnóstico más preciso que en el momento de cumplimentar la ficha diagnóstica.

### Fiabilidad de consistencia interna del MCMI-III y del BDI-II en la muestra de pacientes del estudio

Los índices de fiabilidad de consistencia interna (coeficientes alfa de Cronbach) de las escalas de Depresión Mayor y Distimia del MCMI-III y del BDI-II en la muestra de pacientes del presente estudio están recogidos en la Tabla 7.12.

**Tabla 7.12.** Fiabilidad de las escalas de Depresión Mayor y Distimia del MCMI-III y del BDI-II (alfa de Cronbach)

	Depresión Mayor del MCMI-III	Distimia del MCMI-III	BDI-II
Versión original	0.90 <sup>a</sup>	0.88 <sup>a</sup>	0.93 <sup>b</sup>
Adaptación española	0.90 <sup>c</sup>	0.88 <sup>c</sup>	0.89 <sup>d</sup>
Este estudio	0.81	0.79	0.90

*Nota.* <sup>a</sup>Pacientes ambulatorios estadounidenses con trastornos psicológicos (Millon et al., 2007).

<sup>b</sup>Pacientes ambulatorios estadounidenses con trastornos psicológicos (Beck, Steer y Brown, 1996).

<sup>c</sup>Pacientes ambulatorios españoles con trastornos psicológicos (Millon et al., 2007).

<sup>d</sup>Pacientes ambulatorios españoles con trastornos psicológicos (Sanz, García Vera, Espinosa, Fortún y Vázquez, 2005)

En el caso del MCMI-III, se obtuvieron índices de consistencia interna para la escala de Depresión Mayor (0.81) y de Distimia (0.79) inferiores a los informados en el manual de la adaptación al español de este inventario (0.90 y 0.88, respectivamente; Millon et al., 2007, adaptación española de V. Cardenal y M. P. Sánchez), pero no obstante aceptables o buenos (Prieto y Muñiz, 2000). La consistencia interna del BDI-II

en este estudio fue alta (alfa de Cronbach = 0.90) y similar a los datos encontrados en estudios anteriores (Beck, Steer y Brown, 1996; Dozois, Dobson y Ahnberg, 1998; Titov et al., 2011).

### Correlación entre las escalas de Depresión Mayor y Distimia del MCMI-III y el BDI-II en la muestra de pacientes del estudio

La correlación entre las escalas CC y D del MCMI-III y el BDI-II (véase la Tabla 7.13) fueron similares a las encontradas en otros estudios. En la adaptación española del MCMI-III, la correlación entre las puntuaciones directas de las escalas de CC y D del MCMI-III fue de 0.86, mientras que en el presente estudio fue de 0.88. En este mismo sentido, las correlaciones entre el BDI-II y las escalas de CC y D del MCMI-III fueron en el presente estudio de 0.67 y 0.62, respectivamente, similares a las encontradas en el estudio de Saulsman (2011), de 0.67 y 0.69, respectivamente.

**Tabla 7.13.** Correlación de Pearson entre las escalas de Distimia y Depresión Mayor del MCMI-III y la puntuación total del BDI-II en el estudio segundo

Instrumentos	BDI-II	PD – D	TB – D	PD – CC	TB – CC
BDI-II	1	0.528*	0.679*	0.515*	0.628*
PD – D	0.528*	1	0.656*	0.881*	0.501*
TB – D	0.679*	0.656*	1	0.526*	0.781*
PD – CC	0.515*	0.881*	0.526*	1	0.653*
TB – CC	0.628*	0.501*	0.781*	0.653*	1

*Nota.* \*La correlación es significativa con  $p < 0,01$  (bilateral). PD – CC = Puntuación directa de la escala de Depresión Mayor del MCMI-II; TB – CC = Tasa base de la escala de Depresión Mayor del MCMI-II; PD – D = Puntuación directa de la escala de Distimia del MCMI-II; TB – D = Tasa base de la escala de Distimia del MCMI-II; BDI-II = Puntuación directa del Inventario de Depresión de Beck-II.

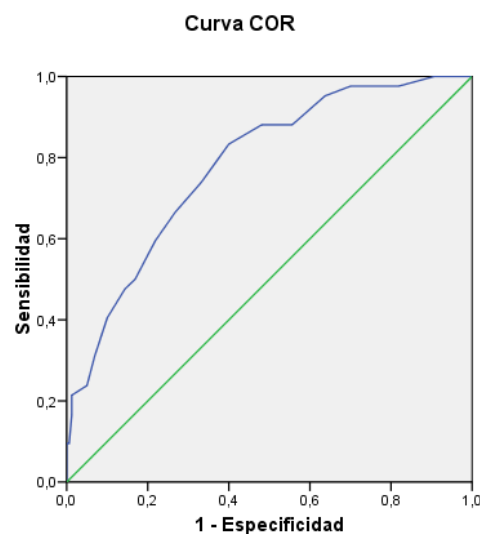


**Eficacia diagnóstica de la escalas CC y D del MCMI-III y del BDI-II para discriminar entre pacientes con diagnóstico de trastorno depresivo mayor y pacientes con otro tipo de diagnóstico según el DSM-IV**

Las curvas ROC realizadas sobre cada una de las puntuaciones directas y de tasa base (TB) de las escalas de CC y D del MCMI-III, así como sobre la puntuación total del BDI-II, en relación con los diagnósticos de trastorno depresivo mayor realizados por los clínicos y recogidos en la ficha diagnóstica, se recogen en las Figuras 7.1 a 7.5.

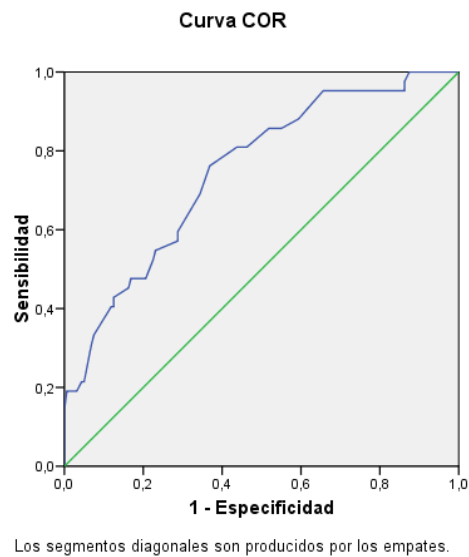
Un análisis visual de las Figuras 7.1 a 7.5 sugería que, en relación con el diagnóstico del trastorno depresivo mayor, las curvas ROC de las escalas CC y D del MCMI-III, tanto para sus puntuaciones directas como de TB, y la curva ROC del BDI-II se separaban moderadamente de la diagonal que reflejaba un diagnóstico basado en el azar.

**Figura 7.1.** Curva ROC de las puntuaciones directas de la escala CC del MCMI-III para el diagnóstico de trastorno depresivo mayor (diagnóstico de la ficha clínica)

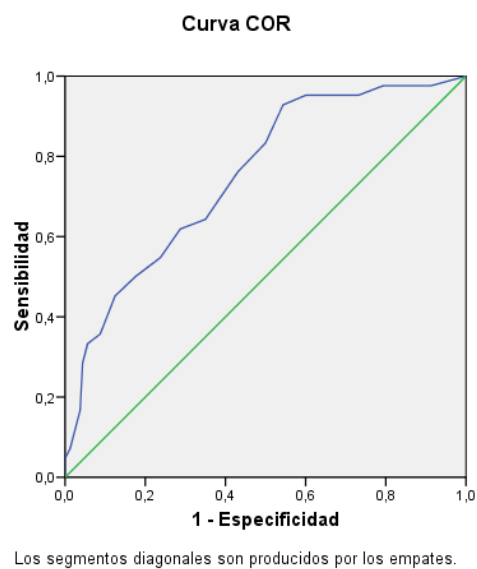


Los segmentos diagonales son producidos por los empates.

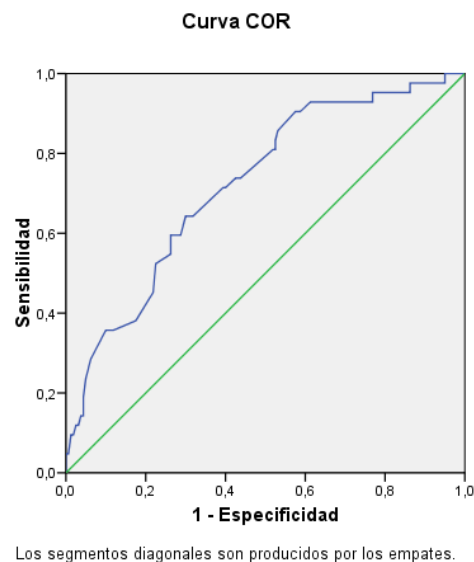
**Figura 7.2.** Curva ROC de las puntuaciones TB de la escala CC del MCMI-III con respecto al diagnóstico de trastorno depresivo mayor (diagnóstico de la ficha clínica)



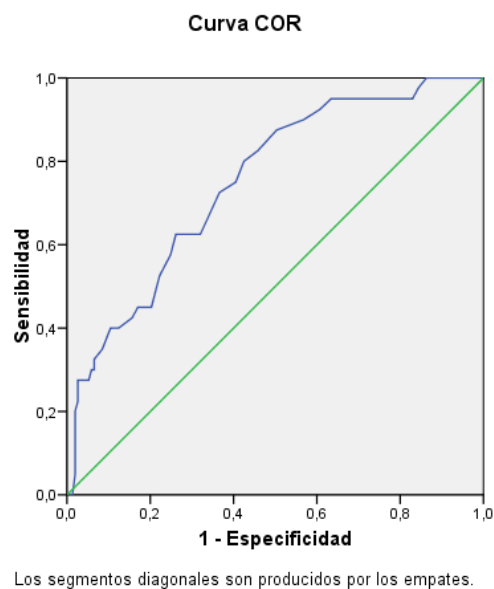
**Figura 7.3.** Curva ROC de las puntuaciones directas de la escala de D del MCMI-III con respecto al diagnóstico de trastorno depresivo mayor (diagnóstico de la ficha clínica)



**Figura 7.4.** Curva ROC de las puntuaciones tasa base de la escala D del MCMI-III con respecto al diagnóstico de trastorno depresivo mayor (diagnóstico de la ficha clínica)



**Figura 7.5.** Curva ROC de las puntuaciones del BDI-III con respecto al diagnóstico de trastorno depresivo mayor (diagnóstico de la ficha clínica)



Todas las escalas tuvieron un área bajo la curva ROC con significación estadística con  $p < 0.001$  (véase la Tabla 7.14). Además, todas las áreas de estas escalas tenían un valor superior a 0.70, por lo que alcanzaron los criterios mínimos recomendados para un instrumento diagnóstico (Swets, 1988), aunque solo las puntuaciones directas y tasa base de la escala de Depresión Mayor (CC) alcanzaban el extremo inferior de los valores recomendados para un instrumento diagnóstico por otros autores ( $> 0.75$ , según Fan et al., 2006).

Igualmente, aunque tanto el análisis visual de las curvas ROC como los mayores valores del área de la curva ROC parecían indicar que las puntuaciones TB y directas de escala CC del MCMI-III presentaban un mejor rendimiento diagnóstico que las puntuaciones directas o TB de la escala de D del MCMI-III, existía un solapamiento de los intervalos de confianza al 95% de las áreas de las curvas ROC correspondientes (véase la Tabla 7.14).

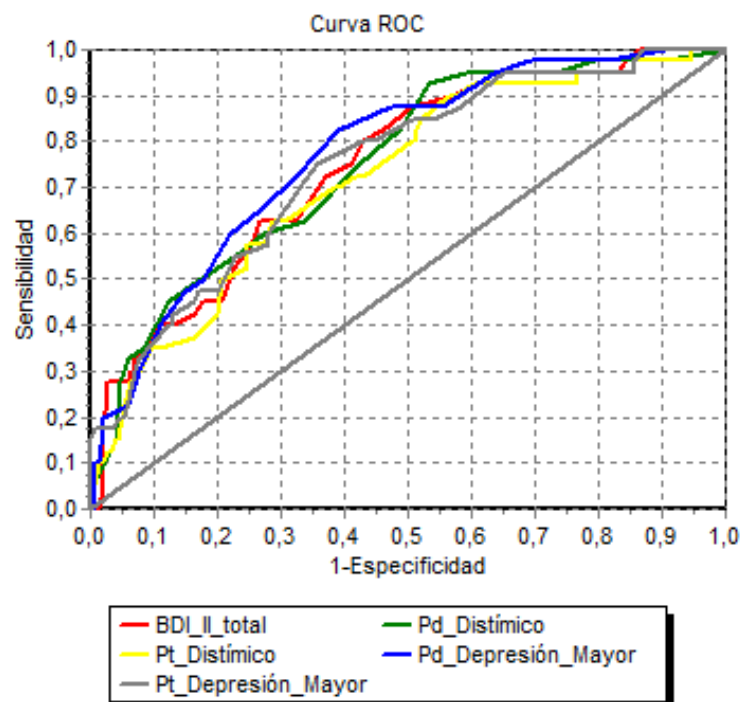
**Tabla 7.14.** Áreas bajo la curva ROC en relación con el diagnóstico de trastorno depresivo mayor para las puntuaciones directas y tasa base de las escalas de Depresión Mayor y Distimia del MCMI-III y de las puntuaciones del BDI-II

Puntuación — Escala	Área bajo ROC para diagnóstico depresión mayor	Intervalo de confianza asintótico del área al 95%	
		Límite inferior	Límite superior
<b>PD — CC</b>	0.767*	0.691	0.843
<b>TB — CC</b>	0.764*	0.687	0.842
<b>PD — D</b>	0.746*	0.664	0.828
<b>TB — D</b>	0.737*	0.656	0.819
<b>BDI-II</b>	0.748*	0.667	0.829

*Nota.* \*Estadísticamente significativas con  $p < 0,001$ . PD — CC = Puntuación directa de la escala de Depresión Mayor del MCMI-III; TB — CC = Tasa base de la escala de Depresión Mayor del MCMI-III; PD — D = Puntuación directa de la escala de Distimia del MCMI-III; TB — D = Tasa base de la escala de Distimia del MCMI-III; BDI-II = Puntuación del Inventario de Depresión de Beck-II.

Con el objetivo de comprobar si efectivamente no había diferencias significativas entre las diferentes curvas ROC de las escalas del MCMI-III y del BDI-II, se realizó un análisis de comparación de curvas ROC. En la Figura 7.6 se puede ver que entre las puntuaciones directas y tasa base de las escalas CC y D del MCMI-III y las puntuaciones totales del BDI-II para el criterio trastorno depresivo mayor existe un alto grado de solapamiento. Efectivamente, las diferencias entre las curvas ROC de las escalas de CC y D del MCMI-III y la puntuación total del BDI-II para el diagnóstico de trastorno depresivo mayor no alcanzaron significación estadística (véase la Tabla 7.15 y el Anexo 7.3).

**Figura 7.6.** Curvas ROC de las puntuaciones directas y TB de las escalas de Distimia y Depresión Mayor del MCMI-III y del BDI-II con respecto al diagnóstico de trastorno depresivo mayor (diagnóstico de la ficha clínica)



**Tabla 7.15.** Comparación de las áreas bajo la curva ROC en relación con el diagnóstico de trastorno depresivo mayor para las puntuaciones directas y TB de las escalas de Depresión Mayor (CC) y Distimia (D) del MCMI-III y del BDI-II

Puntuación – Escala	TB – CC	PD – D	TB – D	BDI-II
PD – CC	$z = 1.058$ $p = 0.290$	$z = 0.751$ $p = 0.452$	$z = 1.330$ $p = 0.183$	$z = 0.614$ $p = 0.539$
TB – CC		$z = 0.0653$ $p = 0.948$	$z = 0.718$ $p = 0.472$	$z = 0.0073$ $p = 0.994$
PD – D			$z = 1.170$ $p = 0.242$	$z = 0.092$ $p = 0.926$
TB – D				$z = 0.809$ $p = 0.418$

*Nota.* PD – CC = Puntuación directa de la escala de Depresión Mayor del MCMI-II; TB – CC = Tasa base de la escala de Depresión Mayor del MCMI-II; PD – D = Puntuación directa de la escala de Distimia del MCMI-II; TB – D = Tasa base de la escala de Distimia del MCMI-II; BDI-II = Puntuación directa del Inventario de Depresión de Beck-II;  $z$  = Estadístico  $z$ ;  $p$  = Nivel de significación estadística.

#### Puntos de corte del MCMI-III y del BDI-II para discriminar entre pacientes con un diagnóstico de trastorno depresivo mayor y pacientes con otro tipo de diagnóstico según el DSM-IV

Se recogen en las Tablas 7.16, 7.17 y 7.18 las coordenadas de las curvas ROC de las escalas de CC y D del MCMI-III y del BDI-II respecto al diagnóstico de trastorno depresivo mayor de la ficha clínica. Estas coordenadas representan la sensibilidad y especificidad específica de cada una de las puntuaciones (directas y TB) de las escalas.

Como puede apreciarse en la Tabla 7.16, con respecto a las puntuaciones directas en la escala de CC del MCMI-III, sensibilidades superiores a 0.70 para el diagnóstico del trastorno depresivo mayor se alcanzaban con puntuaciones directas y TB en la escala de CC del MCMI-III inferiores a 9 y 63, respectivamente, mientras que especificidades superiores a 0.70 suponían puntuaciones directas superiores a 9 y 63, respectivamente.

**Tabla 7.16.** Coordenadas de las curvas ROC de las puntuaciones directas y tasa base de la escala de Depresión Mayor del MCMI-III con respecto al diagnóstico del conjunto de los trastornos depresivos (diagnóstico de la ficha clínica)

Positivo si la puntuación directa de CC es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad	Positivo si la puntuación tasa base de CC es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad
-1.00	1.000	1.000	-1.00	1.000	1.000
.50	1.000	.907	2.50	1.000	0.932
1.50	.976	.820	6.50	1.000	0.919
2.50	.976	.752	8.50	1.000	0.882
3.50	.976	.702	9.50	1.000	0.870
4.50	.952	.640	11.00	0.976	0.870
5.50	.881	.559	12.50	0.976	0.832
6.50	.881	.484	14.00	0.976	0.826
7.50	.833	.404	17.00	0.976	0.770
8.50	.738	.335	19.50	0.976	0.764
9.50	.667	.273	21.50	0.976	0.758
10.50	.595	.224	23.50	0.976	0.720
11.50	.500	.174	27.00	0.976	0.708
12.50	.476	.149	33.00	0.976	0.665
13.50	.405	.106	37.00	0.976	0.652
14.50	.310	.075	41.50	0.905	0.590
15.50	.238	.056	45.50	0.881	0.547
16.50	.214	.019	47.00	0.881	0.540
17.50	.167	.019	50.00	0.881	0.522
18.50	.095	.012	52.50	0.881	0.516
19.50	.095	.006	54.50	0.833	0.460
20.50	.071	.006	57.00	0.833	0.453
21.50	.048	.006	58.50	0.833	0.447
44.50	.024	.006	59.50	0.833	0.435
68.00	.000	.000	61.00	0.786	0.366
			62.50	0.714	0.342
			63.50	0.619	0.286
			64.50	0.595	0.286
			65.50	0.571	0.230
			66.50	0.548	0.224
			67.50	0.476	0.205
			68.50	0.476	0.168
			69.50	0.452	0.161
			70.50	0.429	0.124
			71.50	0.405	0.124
			72.50	0.405	0.118
			73.50	0.333	0.075
			74.50	0.310	0.068

Positivo si la puntuación directa de CC es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad	Positivo si la puntuación tasa base de CC es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad
			75.50	0.214	0.050
			76.50	0.214	0.043
			77.50	0.190	0.031
			78.50	0.190	0.019
			79.50	0.190	0.012
			80.50	0.190	0.006
			82.00	0.143	0.000
			88.50	0.119	0.000
			94.50	0.071	0.000
			95.50	0.048	0.000
			104.00	0.024	0.000
			113.00	0.000	0.000

*Nota.* <sup>a</sup>El punto de corte menor es la mínima puntuación observada menos 1, mientras que el punto de corte mayor es la máxima puntuación observada más 1. Todos los demás puntos de corte son la media de dos puntuaciones observadas ordenadas y consecutivas. CC = Escala de Depresión Mayor del MCMI-III.

En la Tabla 7.17 se recogen las coordenadas de las curvas ROC de las puntuaciones directas y tasa base de la escala de D del MCMI-III con respecto al diagnóstico de trastorno depresivo mayor, las cuales indican que sensibilidades superiores a 0.70 para detectar el trastorno depresivo mayor se alcanzaban con puntuaciones directas y TB en la escala de D del MCMI-III inferiores a 9 y 69 respectivamente, mientras que especificidades superiores a 0.70 suponían puntuaciones tasa base superiores a 10 y 70, respectivamente.

**Tabla 7.17.** Coordenadas de las curvas ROC de las puntuaciones directa y tasa base de la escala de Distimia del MCMI-III con respecto al diagnóstico de trastorno depresivo mayor (diagnóstico de la ficha clínica)

Positivo si la puntuación directa de D es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad	Positivo si la puntuación tasa base de D es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad
-1.00	1.000	1.000	-1.00	1.000	1.000



Positivo si la puntuación directa de D es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad	Positivo si la puntuación tasa base de D es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad
.50	.976	.913	1.00	1.000	0.950
1.50	.976	.869	2.50	0.976	0.950
2.50	.976	.794	4.00	0.976	0.944
3.50	.952	.731	6.00	0.976	0.932
4.50	.952	.681	7.50	0.976	0.925
5.50	.952	.600	8.50	0.976	0.882
6.50	.929	.544	9.50	0.976	0.876
7.50	.833	.500	12.50	0.976	0.863
8.50	.762	.431	15.50	0.976	0.826
9.50	.643	.350	18.00	0.976	0.814
10.50	.619	.288	21.50	0.976	0.789
11.50	.548	.238	25.00	0.976	0.764
12.50	.476	.175	28.50	0.952	0.764
13.50	.429	.125	30.50	0.952	0.702
14.50	.333	.088	34.50	0.952	0.696
15.50	.310	.056	39.00	0.952	0.627
16.50	.262	.044	42.50	0.952	0.609
17.50	.143	.038	47.50	0.929	0.584
18.50	.048	.013	51.00	0.929	0.578
19.50	.024	.000	52.50	0.929	0.571
45.50	.000	.000	55.00	0.881	0.528
			57.50	0.857	0.522
			58.50	0.833	0.522
			59.50	0.833	0.516
			62.00	0.762	0.435
			64.50	0.762	0.429
			65.50	0.762	0.422
			66.50	0.738	0.398
			67.50	0.738	0.391
			69.00	0.667	0.317
			70.50	0.667	0.298
			71.50	0.595	0.286
			72.50	0.595	0.267
			73.50	0.595	0.261
			74.50	0.548	0.261
			75.50	0.524	0.224
			76.50	0.452	0.217
			77.50	0.381	0.174
			78.50	0.357	0.118
			79.50	0.357	0.099
			80.50	0.310	0.075
			81.50	0.286	0.062
			83.50	0.238	0.050

Positivo si la puntuación directa de D es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad	Positivo si la puntuación tasa base de D es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad
			87.00	0.190	0.043
			89.50	0.167	0.043
			90.50	0.143	0.043
			92.00	0.143	0.037
			93.50	0.119	0.031
			94.50	0.119	0.025
			97.00	0.095	0.019
			99.50	0.095	0.012
			100.50	0.048	0.006
			103.00	0.048	0.000
			105.50	0.024	0.000
			107.00	0.000	0.000

*Nota.* <sup>a</sup>El punto de corte menor es la mínima puntuación observada menos 1, mientras que el punto de corte mayor es la máxima puntuación observada más 1. Todos los demás puntos de corte son la media de dos puntuaciones observadas ordenadas y consecutivas. D = Escala de Distimia del MCMI-III.

Finalmente, en la Tabla 7.18 se recogen las coordenadas de la curva ROC de las puntuaciones del BDI-II, las cuales indicaban que sensibilidades superiores a 0.70 para el diagnóstico de trastorno depresivo mayor se alcanzaban con puntuaciones menores de 23 y especificidades superiores a 0.70 suponían puntuaciones mayores de 24.

**Tabla 7.18.** Coordenadas de las curvas ROC de las puntuaciones del BDI-II con respecto al diagnóstico de trastorno depresivo mayor (diagnóstico de la ficha clínica)

Positivo si la puntuación total del BDI-II es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad
0.00	1.000	1.000
2.00	1.000	.987
3.50	1.000	.974
4.50	1.000	.968
5.50	1.000	.942
6.50	1.000	.903
7.50	1.000	.890
8.50	1.000	.864
9.50	.975	.844
10.50	.950	.832

Positivo si la puntuación total del BDI-II es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad
11.50	.950	.792
12.50	.950	.747
13.50	.950	.682
14.50	.950	.643
15.50	.950	.636
16.50	.925	.610
17.50	.900	.571
18.50	.875	.506
19.50	.825	.461
20.50	.800	.429
21.50	.750	.409
22.50	.725	.370
23.50	.625	.325
24.50	.625	.266
25.50	.575	.253
26.50	.525	.227
27.50	.450	.208
28.50	.450	.175
29.50	.425	.162
30.50	.400	.160
31.50	.400	.110
32.50	.350	.091
33.50	.325	.071
34.50	.300	.071
35.50	.300	.065
37.50	.275	.058
39.50	.275	.052
40.50	.275	.045
41.50	.275	.026
42.50	.225	.026
43.50	.200	.019
44.50	.125	.019
45.50	.100	.019
46.50	.075	.019
48.00	.050	.019
50.00	.000	.013
52.00	.000	.000

*Nota.* <sup>a</sup>El punto de corte menor es la mínima puntuación observada menos 1, mientras que el punto de corte mayor es la máxima puntuación observada más 1. Todos los demás puntos de corte son la media de dos puntuaciones observadas ordenadas y consecutivas.

En las Tablas 7.19, 7.20 y 7.21 además de la sensibilidad y la especificidad, se recoge un mayor número de índices de validez diagnóstica (índice de Youden, poder predictivo positivo, poder predictivo negativo y kappa) en relación con el diagnóstico de trastorno depresivo mayor realizado por el clínico y reflejado en la ficha diagnóstica o clínica, para aquellas puntuaciones de las escalas de CC y D del MCMI-III y del BDI-II que han sido propuestas en los manuales correspondientes como puntos de corte para el diagnóstico de trastorno depresivo mayor (TB 75 y TB 85 del MCMI-II) o para la delimitación de diferentes niveles de gravedad de la sintomatología depresiva (BDI-II = 14, 20 y 28). Así mismo, se presentan esos mismos índices de validez diagnóstica para algunas puntuaciones que se encuentran cercanas a las anteriores.

El punto de corte TB 75 de la escala de CC del MCMI-III (véase la Tabla 7.19) mostraba un débil índice de concordancia (0.29) por lo que se situó por debajo de los valores mínimos de referencia para una concordancia diagnóstica moderada ( $\text{kappa} > 0.40$ ), con una alta especificidad (93.17%), una sensibilidad baja (30.95%) y una eficacia global más elevada (80.3). Respecto al punto de corte TB 85, tuvo un índice de concordancia pobre (0.18), probablemente debido a que está situado con una puntuación demasiado elevada, proporcionando un 100% en especificidad (todos los pacientes sin diagnóstico de trastorno depresivo mayor puntuaron por debajo de TB 85) a costa de una baja sensibilidad (11.9%). Ambos puntos de corte se situaron por debajo del valor mínimo de referencia para los índices de sensibilidad ( $> 70\%$ ), por lo que utilizando cualquiera de estos dos puntos de corte quedarían sin detectar a un alto porcentaje de personas que efectivamente tienen un trastorno depresivo mayor. En el caso de las puntuaciones TB 75 y TB 85 tanto el poder predictivo positivo (54.17% y 100%, respectivamente) como el poder predictivo negativo (83.8% y 83.31%, respectivamente) se situaron por encima del valor mínimo de referencia ( $> 50\%$ ). Otras puntuaciones como TB 61 o TB 62 mostraron el mismo índice kappa que la puntuación TB 75 (0.29) y, por tanto, también se

encontraban en el rango de valores más bajos para una concordancia diagnóstica débil ( $\kappa > 0.20$ ), con un mayor equilibrio entre sensibilidad y especificidad (78.57% y 63.35% respectivamente), un mayor índice de Youden (41.92, por encima del valor mínimo considerado aceptable para un instrumento diagnóstico) y un bajo poder predictivo positivo (35.87%).

Las puntuaciones directas 9 y 14 de la escala de CC tuvieron valores de concordancia similares a TB 75 ( $\kappa$ s = 0.29 y 0.3 respectivamente). En el caso de la puntuación directa 9, se encontró una mayor sensibilidad (73.81%) que especificidad (66.46%) y un valor aceptable para el índice de Youden (40.27); mientras que la puntuación directa 11 mostró una mayor especificidad (89.44%) que sensibilidad (38.1%). Ambos puntos de corte tuvieron un bajo poder predictivo positivo (36.47% y 48.48% respectivamente, de nuevo por debajo del 50%, valor mínimo recomendado para un instrumento diagnóstico).

En el caso de la escala D (véase la Tabla 7.20), las puntuaciones TB con mayor índice de concordancia fueron las puntuaciones TB 75 y TB 85, aunque dichos índices fueron también débiles (0.24). Mientras el punto de corte TB 75 tuvo un mayor equilibrio entre sensibilidad y especificidad (54.76% y 73.91% respectivamente), la puntuación TB 85 tuvo un mayor especificidad (95.03%) a costa de una débil sensibilidad (23.81%). Como se ha comentado con respecto a las puntuaciones TB 75 y TB 85 de la escala CC, una sensibilidad inferior al 70% implicaba que un número demasiado elevado de pacientes con diagnóstico de depresión mayor puntuaban por debajo del punto de corte. Ambas puntuaciones TB (75 y 85) de la escala D mostraron un poder predictivo negativo superior a 80% y un menor poder predictivo positivo, ligeramente superior al 50% en el caso de la puntuación TB 85, inferior a ese porcentaje para la puntuación TB 75. Con el mismo índice  $\kappa$  que estas puntuaciones, la puntuación TB 68 tuvo un mayor índice

de Youden, 34.68, con un mayor equilibrio entre sensibilidad y especificidad (73.81% y 60.87%, respectivamente). En el caso de la puntuación directa 15 de la escala D del MCMI-III, se encontraron datos muy similares (aunque con menor sensibilidad, 33.33%) a los comentados para la puntuación TB 75, con un índice kappa ligeramente superior (0.28).

Con respecto al BDI-II (véase la Tabla 7.21), la puntuación total 25 tuvo el mayor índice kappa (0.29), lo que indicaba una concordancia débil (Altman, 1991). Este punto de corte mostraba una mayor especificidad (73.38%) que sensibilidad (62.5%). En cambio, se puede seleccionar el punto de corte 21 ( $\text{kappa} = 0.24$ ) si se quiere primar la sensibilidad (80%) por encima de la especificidad (57.14%), para que el instrumento cumpla una de sus propiedades originales, la de servir como instrumento de *screening*. Además, es precisamente el punto de corte 21 el que mostraba un mayor índice de Youden (37.14). En cambio, el punto de corte 32 tuvo una alta especificidad (88.96%) a costa de una baja sensibilidad (40), con un índice kappa también débil (0.31).

En resumen, los datos recogidos en las Tablas 7.19, 7.20 y 7.21, sugieren que, para los distintos puntos de corte de las escalas CC y D del MCMI-III y del BDI-II, los índices kappa se situaban en el rango bajo de los estándares para una concordancia débil (0.21-0.40) y ninguna puntuación obtuvo un adecuado equilibrio entre sensibilidad y especificidad, aunque había puntuaciones con sensibilidades adecuadas o con especificidades adecuadas.

**Tabla 7.19.** Índices de validez diagnóstica de distintas puntuaciones directas (PD) o tasa base (TB) de la escala de Depresión Mayor (CC) del MCMI-III con respecto al diagnóstico de trastorno depresivo mayor (diagnóstico de la ficha clínica)

Punto de corte	Sensibilidad	Especificidad	Índice de Youden	Poder predictivo positivo	Poder predictivo negativo	Tasa de falsos positivos	Tasa de falsos negativos	Kappa	Eficacia global
<b>TB 75</b>	<b>30.95</b>	<b>93.17</b>	<b>24.12</b>	<b>54.17</b>	<b>83.8</b>	<b>6.83</b>	<b>69.05</b>	<b>0.29</b>	<b>80.3</b>
<b>TB 85</b>	<b>11.9</b>	<b>100</b>	<b>11.9</b>	<b>100</b>	<b>81.31</b>	<b>0</b>	<b>88.1</b>	<b>0.18</b>	<b>81.77</b>
TB 66	57.14	77.02	34.16	39.34	87.32	22.98	42.86	0.29	72.91
TB 67	54.76	77.64	32.4	38.98	86.81	22.36	45.24	0.28	72.91
TB 68	47.62	79.5	27.12	37.74	85.33	20.5	52.38	0.25	72.91
<b>TB 62</b>	<b>78.57</b>	<b>63.35</b>	<b>41.92</b>	<b>35.87</b>	<b>91.89</b>	<b>36.65</b>	<b>21.43</b>	<b>0.29</b>	<b>66.5</b>
<b>TB 61</b>	<b>78.57</b>	<b>63.35</b>	<b>41.92</b>	<b>35.87</b>	<b>91.89</b>	<b>36.65</b>	<b>21.43</b>	<b>0.29</b>	<b>66.5</b>
TB 60	83.33	56.52	39.85	33.33	92.86	43.48	16.67	0.26	62.07
PD 7	88.1	51.55	39.65	32.17	94.32	48.45	11.9	0.24	59.11
<b>PD 9</b>	<b>73.81</b>	<b>66.46</b>	<b>40.27</b>	<b>36.47</b>	<b>90.68</b>	<b>33.54</b>	<b>26.19</b>	<b>0.29</b>	<b>67.98</b>
PD 8	83.33	59.63	42.96	35	93.2	40.37	16.67	0.28	64.53
PD 10	64.29	72.67	36.96	38.03	88.64	27.33	35.71	0.29	70.94
PD 11	57.14	77.64	34.78	40	87.41	22.36	42.86	0.3	73.4
PD 12	47.62	82.61	30.23	41.67	85.81	17.39	52.38	0.29	75.37
PD 13	45.24	85.09	30.33	44.19	85.63	14.91	54.76	0.3	76.85
<b>PD 14</b>	<b>38.1</b>	<b>89.44</b>	<b>27.54</b>	<b>48.48</b>	<b>84.71</b>	<b>10.56</b>	<b>61.9</b>	<b>0.3</b>	<b>78.82</b>
PD 15	28.57	92.55	21.12	50	83.24	7.45	71.43	0.25	79.31
PD 16	21.43	94.41	15.84	50	82.16	5.59	78.57	0.2	79.31
PD 17	19.05	98.14	17.19	72.73	82.29	1.86	80.95	0.24	81.77
PD 18	14.29	98.14	12.43	66.67	81.44	1.86	85.71	0.18	80.79
PD 19	7.14	98.76	5.9	60	80.3	1.24	92.86	0.09	79.8
PD 20	7.14	99.38	6.52	75	80.4	0.62	92.86	0.1	80.3
PD 23	0	99.38	-0.62	0	79.21	0.62	100	-0.01	78.82
PD 28	0	99.38	-0.62	0	79.21	0.62	100	-0.01	78.82
PD 32	0	99.38	-0.62	0	79.21	0.62	100	-0.01	78.82
PD 35	0	99.38	-0.62	0	79.21	0.62	100	-0.01	78.82

**Tabla 7.20** Índices de validez diagnóstica de distintas puntuaciones directas (PD) o tasa base (TB) de la escala de Depresión Mayor (CC) del MCMI-III con respecto al diagnóstico de trastorno depresivo mayor (diagnóstico de la ficha clínica)

Punto de corte	Sensibilidad	Especificidad	Índice de Youden	Poder predictivo positivo	Poder predictivo negativo	Tasa de falsos positivos	Tasa de falsos negativos	Kappa	Eficacia global
<b>TB 85</b>	<b>23.81</b>	<b>95.03</b>	<b>18.84</b>	<b>55.56</b>	<b>82.7</b>	<b>4.97</b>	<b>76.19</b>	<b>0.24</b>	<b>80.3</b>
<b>TB 75</b>	<b>54.76</b>	<b>73.91</b>	<b>28.67</b>	<b>35.38</b>	<b>86.23</b>	<b>26.09</b>	<b>45.24</b>	<b>0.24</b>	<b>69.95</b>
TB 66	76.19	57.76	33.95	32	90.29	42.24	23.81	0.22	61.58
TB 67	73.81	60.25	34.06	32.63	89.81	39.75	26.19	0.23	63.05
<b>TB 68</b>	<b>73.81</b>	<b>60.87</b>	<b>34.68</b>	<b>32.98</b>	<b>89.91</b>	<b>39.13</b>	<b>26.19</b>	<b>0.24</b>	<b>63.55</b>
TB 60	83.33	48.45	31.78	29.66	91.76	51.55	16.67	0.19	55.67
TB 61	76.19	56.52	32.71	31.37	90.1	43.48	23.81	0.21	60.59
TB 62	76.19	56.52	32.71	31.37	90.1	43.48	23.81	0.21	60.59
PD 8	83.33	50.31	33.64	30.43	92.05	49.69	16.67	0.2	57.14
PD 9	76.19	57.14	33.33	31.68	90.2	42.86	23.81	0.22	61.08
PD 10	64.29	65.22	29.51	32.53	87.5	34.78	35.71	0.22	65.02
PD 11	61.9	71.43	33.33	36.11	87.79	28.57	38.1	0.26	69.46
<b>PD 15</b>	<b>33.33</b>	<b>91.3</b>	<b>24.63</b>	<b>50</b>	<b>84</b>	<b>8.7</b>	<b>66.67</b>	<b>0.28</b>	<b>79.31</b>
PD 18	14.29	96.27	10.56	50	81.15	3.73	85.71	0.14	79.31
PD 20	2.38	100	2.38	100	79.7	0	97.62	0.04	79.8
PD 23	0	100	0	100	79.31	0	100	0	79.31
PD 28	0	100	0	100	79.31	0	100	0	79.31
PD 32	0	100	0	100	79.31	0	100	0	79.31
PD 35	0	100	0	100	79.31	0	100	0	79.31



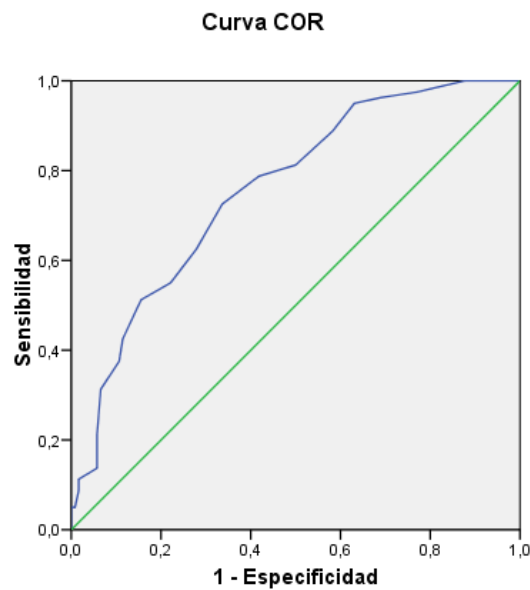
**Tabla 7.21.** Índices de validez diagnóstica de distintas puntuaciones del BDI-II con respecto al diagnóstico de trastorno depresivo mayor (diagnóstico de la ficha clínica)

Punto de corte	Sensibilidad	Especificidad	Índice de Youden	Poder predictivo positivo	Poder predictivo negativo	Tasa de falsos positivos	Tasa de falsos negativos	Kappa	Eficacia global
14	95	31.82	26.82	26.57	96.08	68.18	5	0.14	44.85
15	95	35.71	30.71	27.74	96.49	64.29	5	0.16	47.94
16	95	36.36	31.36	27.94	96.55	63.64	5	0.17	48.45
17	92.5	38.96	31.46	28.24	95.24	61.04	7.5	0.17	50
18	90	42.86	32.86	29.03	94.29	57.14	10	0.18	52.58
19	87.5	49.35	36.85	30.97	93.83	50.65	12.5	0.22	57.22
20	82.5	53.9	36.4	31.73	92.22	46.1	17.5	0.23	59.79
<b>21</b>	<b>80</b>	<b>57.14</b>	<b>37.14</b>	<b>32.65</b>	<b>91.67</b>	<b>42.86</b>	<b>20</b>	<b>0.24</b>	<b>61.86</b>
22	75	59.09	34.09	32.26	90.1	40.91	25	0.23	62.37
23	72.5	62.99	35.49	33.72	89.81	37.01	27.5	0.25	64.95
24	62.5	67.53	30.03	33.33	87.39	32.47	37.5	0.23	66.49
<b>25</b>	<b>62.5</b>	<b>73.38</b>	<b>35.88</b>	<b>37.88</b>	<b>88.28</b>	<b>26.62</b>	<b>37.5</b>	<b>0.29</b>	<b>71.13</b>
27	52.5	77.27	29.77	37.5	86.23	22.73	47.5	0.26	72.16
28	45	79.22	24.22	36	84.72	20.78	55	0.22	72.16
29	45	82.47	27.47	40	85.23	17.53	55	0.26	74.74
30	42.5	83.77	26.27	40.48	84.87	16.23	57.5	0.26	75.26
31	40	87.01	27.01	44.44	84.81	12.99	60	0.28	77.32
<b>32</b>	<b>40</b>	<b>88.96</b>	<b>28.96</b>	<b>48.48</b>	<b>85.09</b>	<b>11.04</b>	<b>60</b>	<b>0.31</b>	<b>78.87</b>
33	35	90.91	25.91	50	84.34	9.09	65	0.29	79.38
34	32.5	92.86	25.36	54.17	84.12	7.14	67.5	0.3	80.41
35	30	92.86	22.86	52.17	83.63	7.14	70	0.27	79.9

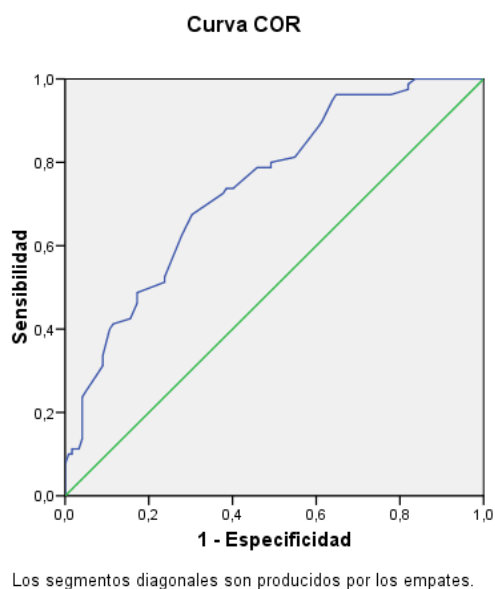
**Eficacia diagnóstica del MCMI-III y del BDI-II para discriminar entre pacientes con cualquier diagnóstico de trastorno depresivo y pacientes con otro tipo de diagnóstico según el DSM-IV**

Las curvas ROC realizadas sobre las puntuaciones directas y de TB de las escalas CC y D del MCMI-III y sobre las puntuaciones del BDI-II, en relación con el diagnóstico de cualquier trastorno depresivo (depresión mayor, distimia, trastorno adaptativo con estado de ánimo deprimido o mixto, duelo y trastorno depresivo no especificado) realizado por los clínicos, se muestran en las Figuras 7.7 a 7.11.

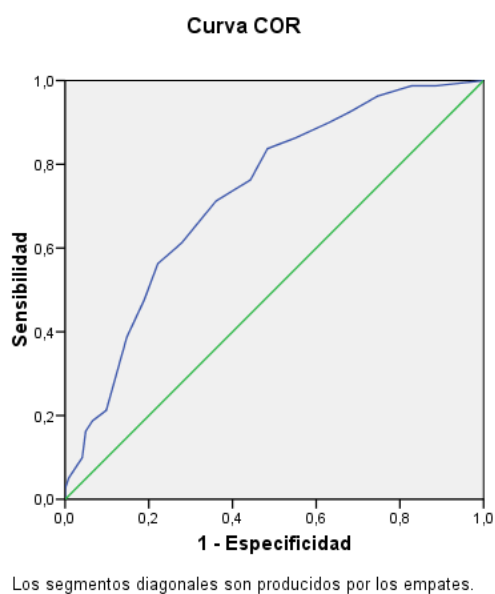
**Figura 7.7.** Curva ROC de las puntuaciones directas de la escala de CC del MCMI-III con respecto al diagnóstico de cualquier trastorno depresivo (diagnóstico de la ficha clínica).



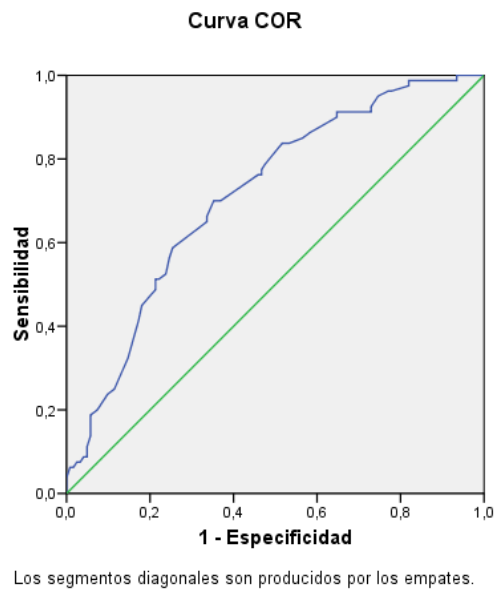
**Figura 7.8.** Curva ROC de las puntuaciones tasa base de la escala de CC del MCMI-III con respecto al diagnóstico cualquier trastorno depresivo (diagnóstico de la ficha clínica).



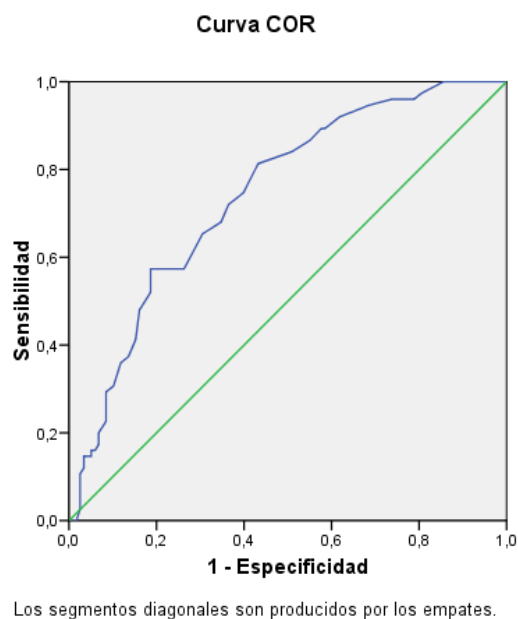
**Figura 7.9.** Curva ROC de las puntuaciones directas de la escala de D del MCMI-III con respecto al diagnóstico de cualquier trastorno depresivo (diagnóstico de la ficha clínica).



**Figura 7.10.** Curva ROC de las puntuaciones tasa base de la escala de D del MCMI-III con respecto al diagnóstico de cualquier trastorno depresivo (diagnóstico de la ficha clínica).



**Figura 7.11.** Curva ROC de las puntuaciones del BDI-II con respecto al diagnóstico de cualquier trastorno depresivo (diagnóstico de la ficha clínica)



Todas las escalas tuvieron un área bajo la curva ROC con significación estadística con  $p < 0.001$  (véase la Tabla 7.22). Además, todas las áreas bajo la curva ROC de estas escalas mostraron valores superiores a 0.70, por lo que alcanzaron los criterios mínimos recomendados para un instrumento diagnóstico (Swets, 1988), aunque solo la puntuación TB de la escala de Depresión Mayor el área bajo su curva ROC alcanzó el extremo inferior de los valores recomendados para un instrumento diagnóstico por otros autores ( $> 0.75$ , según Fan et al., 2006).

**Tabla 7.22.** Áreas bajo la curva ROC para el diagnóstico de cualquier trastorno depresivo realizado por los clínicos para las puntuaciones directas y tasa base de las escalas de Depresión Mayor y Distimia del MCMI-III y del BDI-II

Puntuación – Escala	Área bajo ROC para diagnóstico trastornos depresivos	Intervalo de confianza asintótico del área al 95%	
		Límite inferior	Límite superior
<b>PD – CC</b>	0.745*	0.676	0.814
<b>TB – CC</b>	0.752*	0.683	0.820
<b>PD – D</b>	0.721*	0.649	0.792
<b>TB – D</b>	0.714*	0.642	0.786
<b>BDI-II</b>	0.737*	0.667	0.807

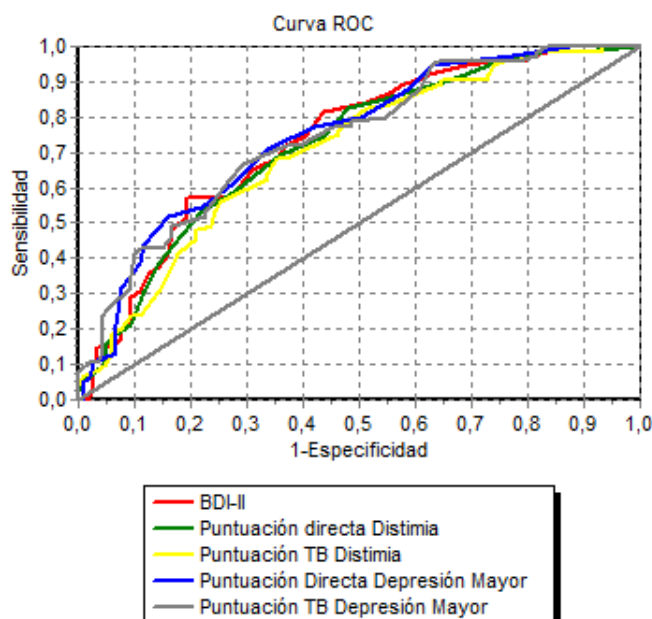
*Nota.* \*Estadísticamente significativas con  $p < 0,001$ . PD – CC = Puntuación directa de la escala de Depresión Mayor del MCMI-III; TB – CC = Tasa base de la escala de Depresión Mayor del MCMI-III; PD – D = Puntuación directa de la escala de Distimia del MCMI-III; TB – D = Tasa base de la escala de Distimia del MCMI-III; BDI-II = Puntuación del Inventario de Depresión de Beck-II.

Igualmente, aunque tanto el análisis visual de las curvas ROC como los mayores valores del área de la curva ROC parecían indicar que las puntuaciones tasa base de la escala CC del MCMI-III presentaban un mejor rendimiento diagnóstico que las puntuaciones directa o TB de la escala D del MCMI-II, las puntuaciones directas de la escala C del MCMI-III o las puntuaciones del BDI-II, existía un solapamiento de los

intervalos de confianza al 95% de las áreas de las curvas ROC correspondientes que no confirmaba dicho análisis visual (véase la Tabla 7.22).

Con el objetivo de comprobar si efectivamente no había diferencias significativas entre las diferentes curvas ROC de las escalas del MCMI-III y del BDI-II, se realizó un análisis de comparación de curvas ROC. En el Figura 7.12 se puede ver las curvas ROC de las puntuaciones de las escalas CC y D del MCMI-III y las puntuaciones totales del BDI-II para el criterio de cualquier diagnóstico de trastorno depresivo según la ficha clínica y, en la Tabla 7.23, se pueden consultar los estadísticos z de comparación por pares de dichas curvas (véase también el anexo 7.4). Efectivamente, tanto el solapamiento de curvas de la Figura 7.12 como los resultados de las pruebas estadísticas revelaron que las diferencias entre las escalas de CC y D del MCMI-III y la puntuación total del BDI-II para el diagnóstico de trastorno depresivo mayor no tenían significación estadística.

**Figura 7.12.** Curvas ROC de las puntuaciones directas y TB de la escala de Distimia y Depresión Mayor del MCMI-III con respecto cualquier trastorno depresivo (diagnóstico de la ficha clínica)



**Tabla 7.23.** Comparación de las áreas bajo la curva ROC en relación con el conjunto de los trastornos depresivos para las puntuaciones directas y tasa base de las escalas de Depresión Mayor y Distimia del MCMI-III y del BDI-II

Puntuación – Escala	TB – CC	PD – D	TB – D	BDI-II
PD – CC	$z = 0.333$	$z = 0.972$	$z = 1.388$	$z = 0.327$
	$p = 0.739$	$p = 0.331$	$p = 0.165$	$p = 0.743$
TB – CC		$z = 0.683$	$z = 1.288$	$z = 0.157$
		$p = 0.494$	$p = 0.197$	$p = 0.875$
PD – D			$z = 1.162$	$z = 0.445$
			$p = 0.245$	$p = 0.656$
TB – D				$z = 0.939$
				$p = 0.347$

*Nota.* PD – CC = Puntuación directa de la escala de Depresión Mayor del MCMI-II; TB – CC = Tasa base de la escala de Depresión Mayor del MCMI-II; PD – D = Puntuación directa de la escala de Distimia del MCMI-II; TB – D = Tasa base de la escala de Distimia del MCMI-II; BDI-II = Puntuación directa del Inventario de Depresión de Beck-II;  $z$  = Estadístico  $z$ ;  $p$  = Nivel de significación estadística.

#### **Puntos de corte del MCMI-III y del BDI-II para discriminar entre pacientes con un diagnóstico del conjunto de los trastornos depresivos y pacientes con otro tipo de diagnóstico según el DSM-IV**

En las Tablas 7.24, 7.25 y 7.26 se recogen las coordenadas de las curvas ROC de las escalas CC y D del MCMI-III y del BDI-II respecto al diagnóstico de la ficha clínica acerca del conjunto de los trastornos depresivos. Estas coordenadas representan la sensibilidad y especificidad específica de cada una de las puntuaciones (directas y TB) de las escalas. Como puede apreciarse en la Tabla 7.24, sensibilidades superiores a 0.70 se alcanzaban para el conjunto de los trastornos depresivos con puntuaciones directas y TB en la escala de CC del MCMI-III inferiores a 8 y 61, respectivamente, mientras que especificidades superiores a 0.70 suponían puntuaciones directas y TB en la escala de CC del MCMI-III superiores a 8 y 62, respectivamente.

**Tabla 7.24.** Coordenadas de las curvas ROC de las puntuaciones directas y tasa base de la escala de Depresión Mayor del MCMI-III con respecto al diagnóstico del conjunto de los trastornos depresivos (diagnóstico de la ficha clínica)

Positivo si la puntuación directa de CC es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad	Positivo si la puntuación tasa base de CC es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad
-1.00	1.000	1.000	-1.00	1.000	1.000
.50	1.000	.878	2.50	1.000	0.911
1.50	.975	.772	6.50	1.000	0.894
2.50	.963	.691	8.50	1.000	0.846
3.50	.950	.634	9.50	1.000	0.829
4.50	.888	.585	11.00	0.988	0.829
5.50	.813	.504	12.50	0.975	0.789
6.50	.788	.423	14.00	0.975	0.780
7.50	.725	.341	17.00	0.975	0.707
8.50	.625	.285	19.50	0.975	0.699
9.50	.538	.228	21.50	0.975	0.691
10.50	.500	.163	23.50	0.975	0.642
11.50	.413	.122	27.00	0.963	0.634
12.50	.363	.114	33.00	0.913	0.610
13.50	.300	.073	37.00	0.900	0.602
14.50	.200	.065	41.50	0.825	0.545
15.50	.125	.065	45.50	0.813	0.488
16.50	.100	.024	47.00	0.800	0.488
17.50	.075	.024	50.00	0.800	0.463
18.50	.038	.016	52.50	0.800	0.455
19.50	.038	.008	54.50	0.750	0.398
20.50	.025	.008	57.00	0.750	0.390
21.50	.013	.008	58.50	0.750	0.382
44.50	.000	.008	59.50	0.738	0.374
68.00	.000	.000	61.00	0.688	0.301
			62.50	0.638	0.276
			63.50	0.538	0.236
			64.50	0.525	0.236
			65.50	0.500	0.171
			66.50	0.475	0.171
			67.50	0.425	0.154
			68.50	0.413	0.114
			69.50	0.400	0.106
			70.50	0.338	0.089
			71.50	0.325	0.089
			72.50	0.313	0.089
			73.50	0.250	0.049
			74.50	0.238	0.041
			75.50	0.150	0.041
			76.50	0.138	0.041
			77.50	0.113	0.033



Positivo si la puntuación directa de CC es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad	Positivo si la puntuación tasa base de CC es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad
			78.50	0.113	0.016
			79.50	0.100	0.016
			80.50	0.100	0.008
			82.00	0.075	0.000
			88.50	0.063	0.000
			94.50	0.038	0.000
			95.50	0.025	0.000
			104.00	0.013	0.000
			113.00	0.000	0.000

*Nota.* <sup>a</sup>El punto de corte menor es la mínima puntuación observada menos 1, mientras que el punto de corte mayor es la máxima puntuación observada más 1. Todos los demás puntos de corte son la media de dos puntuaciones observadas ordenadas y consecutivas. CC = Escala de Depresión Mayor del MCMI-III.

En la Tabla 7.25 se recogen las coordenadas de las curvas ROC de las puntuaciones directas y tasa base de la escala de D del MCMI-III con respecto al diagnóstico del conjunto de los trastornos depresivos, las cuales indican que sensibilidades superiores a 0.70 para detectar el conjunto de los trastornos depresivos se alcanzaban con puntuaciones directas y TB inferiores a 9 y 66, respectivamente, mientras que especificidades superiores a 0.70 suponían puntuaciones directas y de tasa base superiores a 9 y 69, respectivamente.

**Tabla 7.25.** Coordenadas de las curvas ROC de las puntuaciones directas y tasa base de la escala de Distimia del MCMI-III con respecto al diagnóstico del conjunto de los trastornos depresivos (diagnóstico de la ficha clínica)

Positivo si la puntuación directa de D es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad	Positivo si la puntuación tasa base de D es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad
-1.00	1.000	1.000	-1.00	1.000	1.000
0.50	0.988	0.886	1.00	1.000	0.935
1.50	0.988	0.829	2.50	0.988	0.935
2.50	0.963	0.748	4.00	0.988	0.927
3.50	0.925	0.683	6.00	0.988	0.911
4.50	0.900	0.634	7.50	0.988	0.902
5.50	0.863	0.545	8.50	0.988	0.846

Positivo si la puntuación directa de D es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad	Positivo si la puntuación tasa base de D es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad
6.50	0.838	0.480	9.50	0.988	0.837
7.50	0.763	0.439	12.50	0.988	0.821
8.50	0.713	0.358	15.50	0.975	0.780
9.50	0.613	0.276	18.00	0.975	0.764
10.50	0.563	0.220	21.50	0.963	0.740
11.50	0.475	0.187	25.00	0.938	0.724
12.50	0.375	0.146	28.50	0.925	0.724
13.50	0.288	0.122	30.50	0.925	0.642
14.50	0.200	0.098	34.50	0.913	0.642
15.50	0.175	0.065	39.00	0.875	0.577
16.50	0.150	0.049	42.50	0.863	0.561
17.50	0.088	0.041	47.50	0.850	0.528
18.50	0.038	0.008	51.00	0.850	0.520
19.50	0.013	0.000	52.50	0.850	0.512
21.00	0.000	0.000	55.00	0.800	0.472
			57.50	0.788	0.463
			58.50	0.775	0.463
			59.50	0.775	0.455
			62.00	0.713	0.366
			64.50	0.713	0.358
			65.50	0.713	0.350
			66.50	0.675	0.333
			67.50	0.663	0.333
			69.00	0.600	0.252
			70.50	0.575	0.244
			71.50	0.525	0.236
			72.50	0.513	0.220
			73.50	0.513	0.211
			74.50	0.488	0.211
			75.50	0.450	0.179
			76.50	0.413	0.171
			77.50	0.325	0.146
			78.50	0.250	0.114
			79.50	0.238	0.098
			80.50	0.200	0.073
			81.50	0.188	0.057
			83.50	0.138	0.057
			87.00	0.113	0.049
			89.50	0.100	0.049
			90.50	0.088	0.049
			92.00	0.088	0.041
			93.50	0.075	0.033
			94.50	0.075	0.024
			97.00	0.063	0.016
			99.50	0.063	0.008
			100.50	0.038	0.000

Positivo si la puntuación directa de D es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad	Positivo si la puntuación tasa base de D es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad
			103.00	0.025	0.000
			105.50	0.013	0.000
			107.00	0.000	0.000

*Nota.* <sup>a</sup>El punto de corte menor es la mínima puntuación observada menos 1, mientras que el punto de corte mayor es la máxima puntuación observada más 1. Todos los demás puntos de corte son la media de dos puntuaciones observadas ordenadas y consecutivas. D = Escala de Distimia del MCMI-III.

Finalmente, en la Tabla 7.26 se recogen las coordenadas de las curvas ROC de las puntuaciones del BDI-II con respecto al diagnóstico del conjunto de los trastornos depresivos, las cuales indican que sensibilidades superiores a 0.70 para el conjunto de los trastornos depresivos se alcanzaban con puntuaciones en el BDI-II inferiores a 21, mientras que especificidades superiores a 0.70 suponían puntuaciones en el BDI-II superiores a 23.

**Tabla 7.26.** Coordenadas de las curvas ROC de las puntuaciones del BDI-II con respecto al diagnóstico del conjunto de los trastornos depresivos (diagnóstico de la ficha clínica)

Positivo si la puntuación total del BDI-II es mayor o igual que <sup>a</sup>	Sensibilidad	1 – especificidad
.00	1.000	1.000
2.00	1.000	.983
3.50	1.000	.966
4.50	1.000	.958
5.50	1.000	.924
6.50	1.000	.874
7.50	1.000	.857
8.50	.987	.832
9.50	.973	.807
10.50	.960	.790
11.50	.960	.739
12.50	.947	.689
13.50	.920	.622
14.50	.893	.588
15.50	.893	.580
16.50	.867	.555

Positivo si la puntuación total del BDI-II es mayor o igual que <sup>a</sup>	Sensibilidad	1 – especificidad
17.50	.840	.513
18.50	.813	.437
19.50	.747	.403
20.50	.720	.370
21.50	.680	.353
22.50	.653	.311
23.50	.573	.269
24.50	.573	.193
25.50	.520	.193
26.50	.480	.168
27.50	.413	.160
28.50	.373	.143
29.50	.360	.126
30.50	.307	.109
31.50	.293	.092
32.50	.227	.092
33.50	.200	.076
34.50	.187	.076
35.50	.173	.076
37.50	.160	.067
39.50	.160	.059
40.50	.147	.059
41.50	.147	.034
42.50	.120	.034
43.50	.107	.025
44.50	.067	.025
45.50	.053	.025
46.50	.040	.025
48.00	.027	.025
50.00	.000	.017
52.00	.000	.000

*Nota.* <sup>a</sup>El punto de corte menor es la mínima puntuación observada menos 1, mientras que el punto de corte mayor es la máxima puntuación observada más 1. Todos los demás puntos de corte son la media de dos puntuaciones observadas ordenadas y consecutivas.

En las Tablas 7.27, 7.28 y 7.29, además de la sensibilidad y la especificidad, se recoge un mayor número de índices de validez diagnóstica (índice de Youden, poder predictivo positivo, poder predictivo negativo y kappa) en relación con el diagnóstico de cualquier trastorno depresivo realizado por el clínico y reflejado en la ficha clínica, para aquellas puntuaciones de las escalas de CC y D del MCMI-III y del BDI-II que han sido propuestas en los manuales correspondientes como puntos de corte para el

diagnóstico de trastorno depresivo mayor (TB 75 y TB 85 del MCMI-II) o para la delimitación de diferentes niveles de gravedad de la sintomatología depresiva (BDI-II = 14, 20 y 28). Así mismo, se presentan esos mismos índices de validez diagnóstica para algunas puntuaciones que se encuentran cercanas a las anteriores.

Con respecto a la escala CC (véase la Tabla 7.27), se puede ver que las puntuaciones TB 75 y TB 85 no fueron las puntuaciones TB con mayor índice kappa, índice de Youden o índice de eficacia global. Tanto la puntuación TB 75 como la puntuación TB 85 tuvieron una sensibilidad muy inferior al 70% (23.75% y 6.25%, respectivamente), con una especificidad alta, 95.93% y 100% respectivamente, de manera que ambas tuvieron un índice de Youden bajo de 19.68 y 6.25, respectivamente. Ambas también tuvieron un índice kappa débil (0.22 en el caso de la TB 75 y 0.07 en el caso de la TB 85). Otros puntos de corte como la TB 61 y la TB 62 tuvieron índices kappa más elevados, aunque también débiles (0.37), pero con una sensibilidad y especificidad cercanas a 70% (68.75% y 69.67%, respectivamente). Estas puntuaciones TB 81 y TB 62 fueron además las puntuaciones TB de la escala CC con un índice de Youden más elevado para el diagnóstico del conjunto de los trastornos depresivos, 38.42, aunque se situó ligeramente por debajo del valor aceptable para este estadístico (40). La puntuación directa 8 de la escala CC tuvo un índice de Youden ligeramente más elevado (38.89) y el mismo índice kappa (0.37), con una sensibilidad del 72.5% y una especificidad de 66.39%.

En cuanto a la escala D del MCMI-III (véase la Tabla 7.28) y para el diagnóstico del conjunto de los trastornos depresivos, sus puntuaciones TB 75 y TB 85 tuvieron una sensibilidad por debajo del valor recomendado (70%). En concreto, la puntuación TB 75 tuvo un índice kappa débil (0.28), con una sensibilidad también débil de 48.75%, aunque con una especificidad superior al 70% (78.69%). Por su parte, la puntuación TB 85 tuvo un índice de concordancia pobre (0.09), con una muy alta especificidad (94.26%) a costa de una muy baja sensibilidad (13.95). Ambos puntos de corte tenían

un poder predictivo positivo inferior al poder predictivo negativo, superiores en ambos casos al 50%. La puntuación TB 66 tuvo un kappa también débil, pero más elevado (0.35) y una sensibilidad mayor (71.25%), con una especificidad de 65.04% y con el mayor índice de Youden (36.29). Las puntuaciones directas 9 y 11 tuvieron los mayores kappas, aunque de nuevo por debajo de 0.40 (0.34 y 0.35 respectivamente), con un índice de Youden cercano a 35. La puntuación directa 9 tuvo una mayor sensibilidad que la puntuación 11 (71.25% frente a 56.25%) mientras que la puntuación directa 11 tuvo una mayor especificidad (78.05% frente a 64.23%), pero no se encontró ningún punto de corte que simultáneamente mostrara sensibilidades y especificidades mayores de 70%.

Las puntuaciones del BDI-II con mejores índices de eficacia para el diagnóstico del conjunto de los trastornos depresivos (véase la Tabla 7.29) fueron las puntuaciones 19 y 25, si tenemos en cuenta que fueron las que tenían un mayor kappa, aunque débil (0.34 y 0.39 respectivamente), y un mayor índice de Youden (37.63 y 38, respectivamente). La puntuación 19 mostró mayor sensibilidad, (81.33%) que especificidad (56.3%), mientras que la puntuación 25 mostró mayor especificidad (80.67%) que sensibilidad (57.33%).

Observando las tres tablas que presentan los índices de eficacia diagnóstica de diferentes puntuaciones de las escalas CC y D del MCMI-III y del BDI-II para el diagnóstico del conjunto de trastornos depresivos (Tablas 7.27, 7.28 y 7.29), se puede constatar que, al igual que se comentó con respecto al diagnóstico del trastorno depresivo mayor, el poder predictivo positivo de los instrumentos para el diagnóstico del conjunto de los trastornos depresivos es menor que su poder predictivo negativo, aunque el poder predictivo positivo parece, en general, ligeramente superior en el caso del diagnóstico del conjunto de trastornos depresivos que en el caso del diagnóstico del trastorno depresivo mayor.

**Tabla 7.27.** Índices de validez diagnóstica de distintas puntuaciones directas (PD) o tasa base (TB) de la escala de Depresión Mayor (CC) del MCMI-III con respecto al diagnóstico del conjunto de los trastornos depresivos (diagnóstico de la ficha clínica)

Punto de corte	Sensibilidad	Especificidad	Índice Youden	Poder predictivo positivo	Poder predictivo negativo	Tasa de falsos positivos	Tasa de falsos negativos	Kappa	Eficacia global
<b>TB 75</b>	<b>23.75</b>	<b>95.9</b>	<b>19.65</b>	<b>79.17</b>	<b>65.73</b>	<b>4.1</b>	<b>76.25</b>	<b>0.22</b>	<b>67.33</b>
<b>TB 85</b>	<b>6.25</b>	<b>100</b>	<b>6.25</b>	<b>100</b>	<b>61.93</b>	<b>0</b>	<b>93.75</b>	<b>0.07</b>	<b>62.87</b>
TB 66	50	82.79	32.79	65.57	71.63	17.21	50	0.34	69.8
TB 67	47.5	82.79	30.29	64.41	70.63	17.21	52.5	0.32	68.81
TB 68	42.5	84.43	26.93	64.15	69.13	15.57	57.5	0.29	67.82
<b>TB 62</b>	<b>68.75</b>	<b>69.67</b>	<b>38.42</b>	<b>59.78</b>	<b>77.27</b>	<b>30.33</b>	<b>31.25</b>	<b>0.37</b>	<b>69.31</b>
<b>TB 61</b>	<b>68.75</b>	<b>69.67</b>	<b>38.42</b>	<b>59.78</b>	<b>77.27</b>	<b>30.33</b>	<b>31.25</b>	<b>0.37</b>	<b>69.31</b>
TB 60	73.75	62.3	36.05	56.19	78.35	37.7	26.25	0.34	66.83
PD 7	78.75	58.2	36.95	55.26	80.68	41.8	21.25	0.34	66.34
PD 9	62.5	72.13	34.63	59.52	74.58	27.87	37.5	0.34	68.32
<b>PD 8</b>	<b>72.5</b>	<b>66.39</b>	<b>38.89</b>	<b>58.59</b>	<b>78.64</b>	<b>33.61</b>	<b>27.5</b>	<b>0.37</b>	<b>68.81</b>
PD 10	53.75	77.87	31.62	61.43	71.97	22.13	46.25	0.32	68.32
PD 11	50	84.43	34.43	67.8	72.03	15.57	50	0.36	70.79
PD 12	41.25	88.52	29.77	70.21	69.68	11.48	58.75	0.32	69.8
PD 13	36.25	89.34	25.59	69.05	68.13	10.66	63.75	0.28	68.32
PD 14	30	93.44	23.44	75	67.06	6.56	70	0.26	68.32
PD 15	20	94.26	14.26	69.57	64.25	5.74	80	0.16	64.85
PD 16	12.5	94.26	6.76	58.82	62.16	5.74	87.5	0.08	61.88
PD 17	10	98.36	8.36	80	62.5	1.64	90	0.1	63.37
PD 18	23.75	95.9	19.65	79.17	65.73	4.1	76.25	0.22	67.33
PD 19	6.25	100	6.25	100	61.93	0	93.75	0.07	62.87
PD 20	50	82.79	32.79	65.57	71.63	17.21	50	0.34	69.8
PD 23	47.5	82.79	30.29	64.41	70.63	17.21	52.5	0.32	68.81
PD 28	42.5	84.43	26.93	64.15	69.13	15.57	57.5	0.29	67.82
PD 32	68.75	69.67	38.42	59.78	77.27	30.33	31.25	0.37	69.31
PD 35	68.75	69.67	38.42	59.78	77.27	30.33	31.25	0.37	69.31

**Tabla 7.28.** Índices de validez diagnóstica de distintas puntuaciones directas (PD) o tasa base (TB) de la escala de Distimia (D) del MCMI-III con respecto al diagnóstico del conjunto de los trastornos depresivos (diagnóstico de la ficha clínica)

Punto de corte	Sensibilidad	Especificidad	Índice de Youden	Poder predictivo positivo	Poder predictivo negativo	Tasa de falsos positivos	Tasa de falsos negativos	Kappa	Eficacia global
<b>TB 85</b>	<b>13.75</b>	<b>94.26</b>	<b>8.01</b>	<b>61.11</b>	<b>62.5</b>	<b>5.74</b>	<b>86.25</b>	<b>0.09</b>	<b>62.38</b>
<b>TB 75</b>	<b>48.75</b>	<b>78.69</b>	<b>27.44</b>	<b>60</b>	<b>70.07</b>	<b>21.31</b>	<b>51.25</b>	<b>0.28</b>	<b>66.83</b>
<b>TB 66</b>	<b>71.25</b>	<b>64.75</b>	<b>36</b>	<b>57</b>	<b>77.45</b>	<b>35.25</b>	<b>28.75</b>	<b>0.35</b>	<b>67.33</b>
TB 67	67.5	66.39	33.89	56.84	75.7	33.61	32.5	0.33	66.83
TB 68	66.25	66.39	32.64	56.38	75	33.61	33.75	0.32	66.34
TB 60	77.5	54.1	31.6	52.54	78.57	45.9	22.5	0.29	63.37
TB 61	71.25	63.11	34.36	55.88	77	36.89	28.75	0.33	66.34
TB 62	71.25	63.11	34.36	55.88	77	36.89	28.75	0.33	66.34
PD 8	76.25	55.74	31.99	53.04	78.16	44.26	23.75	0.3	63.86
<b>PD 9</b>	<b>71.25</b>	<b>63.93</b>	<b>35.18</b>	<b>56.44</b>	<b>77.23</b>	<b>36.07</b>	<b>28.75</b>	<b>0.34</b>	<b>66.83</b>
PD 10	61.25	72.13	33.38	59.04	73.95	27.87	38.75	0.33	67.82
<b>PD 11</b>	<b>56.25</b>	<b>77.87</b>	<b>34.12</b>	<b>62.5</b>	<b>73.08</b>	<b>22.13</b>	<b>43.75</b>	<b>0.35</b>	<b>69.31</b>
PD 15	20	90.16	10.16	57.14	63.22	9.84	80	0.11	62.38
PD 18	8.75	95.9	4.65	58.33	61.58	4.1	91.25	0.05	61.39
PD 20	1.25	100	1.25	100	60.7	0	98.75	0.02	60.89
PD 23	0	100	0	100	60.4	0	100	0	60.4
PD 28	0	100	0	100	60.4	0	100	0	60.4
PD 32	0	100	0	100	60.4	0	100	0	60.4
PD 35	0	100	0	100	60.4	0	100	0	60.4



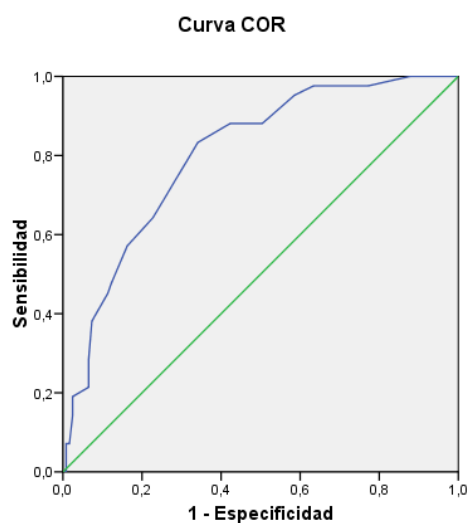
**Tabla 7.29.** Índices de validez diagnóstica de distintas puntuaciones del BDI-II con respecto al diagnóstico del conjunto de los trastornos depresivos (diagnóstico de la ficha clínica)

Punto de corte	Sensibilidad	Especificidad	Índice de Youden	Poder predictivo positivo	Poder predictivo negativo	Tasa de falsos positivos	Tasa de falsos negativos	Kappa	Eficacia global
14	95	32.03	27.03	26.76	96.08	67.97	5	0.14	45.08
15	95	35.95	30.95	27.94	96.49	64.05	5	0.16	48.19
16	95	36.6	31.6	28.15	96.55	63.4	5	0.17	48.7
17	92.5	39.22	31.72	28.46	95.24	60.78	7.5	0.17	50.26
18	90	43.14	33.14	29.27	94.29	56.86	10	0.19	52.85
19	87.5	49.67	37.17	31.25	93.83	50.33	12.5	0.22	57.51
20	82.5	54.25	36.75	32.04	92.22	45.75	17.5	0.23	60.1
<b>21</b>	<b>80</b>	<b>57.52</b>	<b>37.52</b>	<b>32.99</b>	<b>91.67</b>	<b>42.48</b>	<b>20</b>	<b>0.25</b>	<b>62.18</b>
22	75	59.48	34.48	32.61	90.1	40.52	25	0.23	62.69
23	72.5	63.4	35.9	34.12	89.81	36.6	27.5	0.25	65.28
24	62.5	67.97	30.47	33.78	87.39	32.03	37.5	0.23	66.84
<b>25</b>	<b>62.5</b>	<b>73.86</b>	<b>36.36</b>	<b>38.46</b>	<b>88.28</b>	<b>26.14</b>	<b>37.5</b>	<b>0.3</b>	<b>71.5</b>
27	52.5	77.78	30.28	38.18	86.23	22.22	47.5	0.27	72.54
28	45	79.74	24.74	36.73	84.72	20.26	55	0.23	72.54
29	45	83.01	28.01	40.91	85.23	16.99	55	0.27	75.13
30	42.5	84.31	26.81	41.46	84.87	15.69	57.5	0.27	75.65
31	40	87.58	27.58	45.71	84.81	12.42	60	0.29	77.72
32	40	89.54	29.54	50	85.09	10.46	60	0.32	79.27
33	35	91.5	26.5	51.85	84.34	8.5	65	0.3	79.79
34	32.5	93.46	25.96	56.52	84.12	6.54	67.5	0.31	80.83
35	30	93.46	23.46	54.55	83.63	6.54	70	0.28	80.31

**Eficacia diagnóstica de las escalas CC y D del MCMI-III y del BDI-II para discriminar entre pacientes con diagnóstico de trastorno depresivo mayor y pacientes con otro tipo de trastorno mental no depresivo según el DSM-IV**

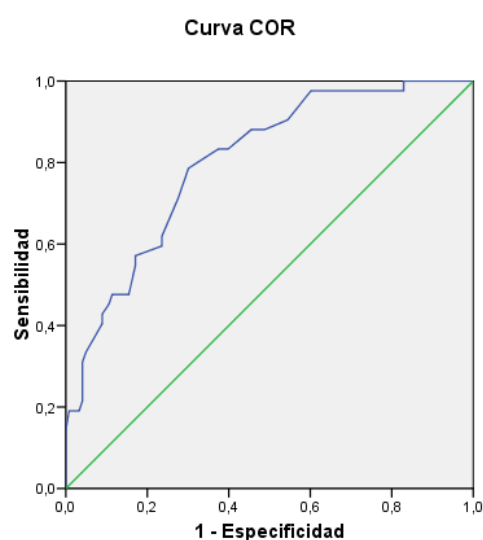
Con el objetivo de comprobar la eficacia diagnóstica de los instrumentos siguiendo las recomendaciones de Ganellen (1996) ya comentadas en el capítulo 6 y en la introducción de este capítulo, se realizó el análisis de las curvas ROC para el diagnóstico del trastorno depresivo mayor excluyendo de la muestra aquellos pacientes que presentaban un diagnóstico del conjunto de trastornos depresivos (distimia, trastorno del estado de ánimo no especificado, duelo o trastorno adaptativo con estado de ánimo depresivo). Las curvas ROC obtenidas para las escalas CC y D del MCMI-III y para el BDI-II respecto al diagnóstico del trastorno depresivo mayor, excluidos los pacientes con el diagnóstico de otros trastornos depresivos, se recogen en las Figuras 7.13 a 7.17.

**Figura 7.13.** Curva ROC de las puntuaciones directas de la escala de Depresión Mayor del MCMI-III con respecto al diagnóstico de trastorno depresivo mayor (pacientes con otros trastornos depresivos excluidos; diagnóstico de la ficha clínica)

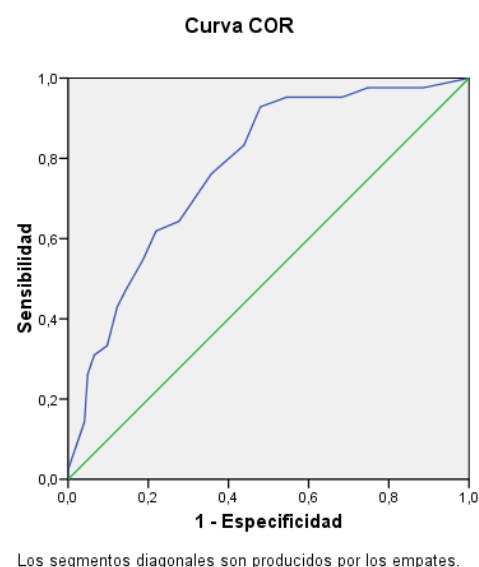


Los segmentos diagonales son producidos por los empates.

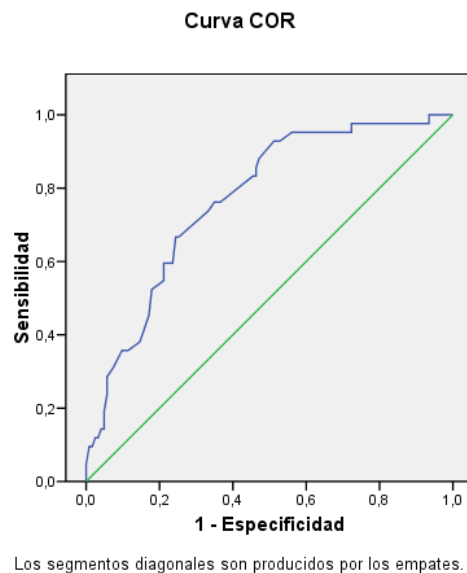
**Figura 7.14.** Curva ROC de las puntuaciones TB de la escala de Depresión Mayor del MCMI-III con respecto al diagnóstico trastorno depresivo mayor (pacientes con otros trastornos depresivos excluidos; diagnóstico de la ficha clínica)



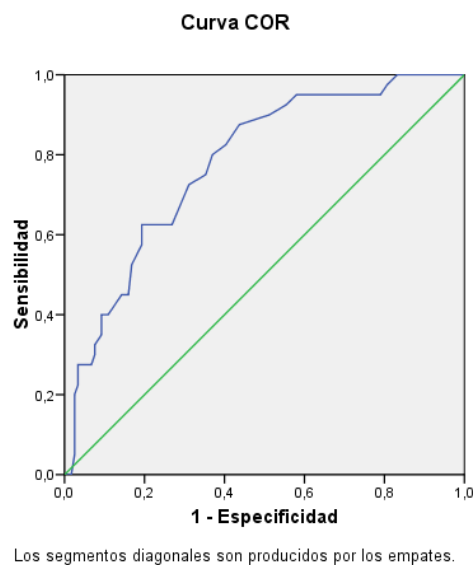
**Figura 7.15.** Curva ROC de las puntuaciones directas de la escala de Distimia del MCMI-III con respecto al diagnóstico de trastorno depresivo mayor (pacientes con otros trastornos depresivos excluidos; diagnóstico de la ficha clínica)



**Figura 7.16.** Curva ROC de las puntuaciones TB de la escala de Distimia del MCMI-III con respecto al diagnóstico de trastorno depresivo mayor (pacientes con otros trastornos depresivos excluidos; diagnóstico de la ficha clínica)



**Figura 7.17.** Curva ROC de las puntuaciones del BDI-II con respecto al diagnóstico de trastorno depresivo mayor (pacientes con otros trastornos depresivos excluidos; diagnóstico de la ficha clínica)



Se puede observar a través de un análisis visual de las Figuras 7.13 a 7.16 que las curvas ROC de las escalas CC y D del MCMI-III se separaban moderadamente de los segmentos diagonales producidos por los empates entre sensibilidad y la tasa de falsos positivos (representados en las gráficas mediante la línea diagonal verde). De hecho, como se puede observar en la Tabla 7.30, el análisis estadístico indicaba que las áreas bajo esas curvas ROC se distanciaban de forma estadísticamente significativa del área de 0.50 que representa un diagnóstico basado en el azar y alcanzaban el criterio mínimo recomendado por algunos investigadores para un instrumento diagnóstico ( $> 0.70$ , según Swets, 1988), así como el criterio más estricto propuesto por otros investigadores ( $> 0.75$ , según Fan et al., 2006).

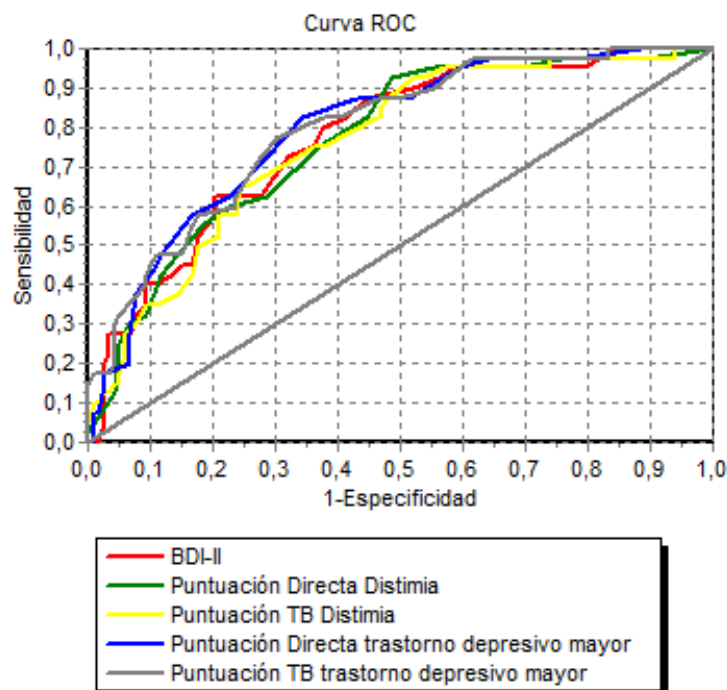
**Tabla 7.30.** Áreas bajo la curva ROC para el diagnóstico de trastorno depresivo mayor (otros trastornos depresivos excluidos) para las puntuaciones directas y tasa base de las escalas de Depresión Mayor y Distimia del MCMI-III y del BDI-II

Puntuación — Escala	Área bajo ROC para diagnóstico depresión mayor	Intervalo de confianza asintótico del área al 95%	
		Límite inferior	Límite superior
PD — CC	0.800*	0.725	0.874
TB — CC	0.801*	0.726	0.876
PD — D	0.772*	0.692	0.851
TB — D	0.764*	0.683	0.844
BDI-II	0.778*	0.701	0.858

*Nota.* \*Estadísticamente significativa con  $p < 0,001$ . PD — CC = Puntuación directa de la escala de Depresión Mayor del MCMI-II; TB — CC = Tasa base de la escala de Depresión Mayor del MCMI-II; PD — D = Puntuación directa de la escala de Distimia del MCMI-II; TB — D = Tasa base de la escala de Distimia del MCMI-II; BDI-II = Puntuación directa del Inventario de Depresión de Beck-II.

El análisis visual de las curvas ROC podría sugerir que los valores de las áreas bajo la curva ROC de las puntuaciones directas o tasa base de la escala CC del MCMI-III eran mayores que los correspondientes de las puntuaciones directas o tasa base de la escala de D del MCMI-II. Sin embargo, existía un solapamiento de los intervalos de confianza al 95% de esas áreas bajo las curvas ROC, lo que no permitía sustanciar empíricamente esa impresión (véase la Tabla 7.30).

**Figura 7.18.** Comparación de las curvas ROC de las puntuaciones directas y TB de la escalas de Distimia y Depresión Mayor del MCMI-III y de las puntuaciones del BDI-II con respecto al diagnóstico de trastorno depresivo mayor (pacientes con otros trastornos depresivos excluidos)



Con el objetivo de comprobar más fehacientemente si existían diferencias significativas entre las diferentes curvas ROC de las escalas del MCMI-III y del BDI-II, se

realizó un análisis de comparación de curvas ROC. En el Figura 7.18 se puede observar que las curvas de las puntuaciones directas y TB de las escalas CC y D del MCMI-III y de las puntuaciones del BDI-II para el diagnóstico del trastorno depresivo mayor excluidos los pacientes con otros trastornos depresivos, prácticamente se solapaban. Es más, en la Tabla 7.31 se puede comprobar que, efectivamente, según los resultados de los estadísticos  $z$  de comparación de pares de áreas bajo la curva ROC, ninguna de las diferencias entre las escalas CC y D del MCMI-III o entre éstas y el BDI-II era estadísticamente significativa (véase también el Anexo 7.5).

**Tabla 7.31.** Comparación de las áreas bajo la curva ROC en relación con el diagnóstico de trastorno depresivo mayor (pacientes con otros trastornos depresivos excluidos) para las puntuaciones directas y tasa base de las escalas de Depresión Mayor y Distimia del MCMI-III y para las puntuaciones del BDI-II

Puntuación – Escala	TB – CC	PD – D	TB – D	BDI-II
PD – CC	$z = 0.108$	$z = 0.898$	$z = 1.053$	$z = 0.543$
	$p = 0.913$	$p = 0.369$	$p = 0.292$	$p = 0.586$
TB – CC		$z = 0.855$	$z = 1.024$	$z = 0.541$
		$p = 0.392$	$p = 0.305$	$p = 0.588$
PD – D			$z = 0.635$	$z = 0.201$
			$p = 0.525$	$p = 0.840$
TB – D				$z = 0.426$
				$p = 0.669$

*Nota.* PD – CC = Puntuación directa de la escala de Depresión Mayor del MCMI-II; TB – CC = Tasa base de la escala de Depresión Mayor del MCMI-II; PD – D = Puntuación directa de la escala de Distimia del MCMI-II; TB – D = Tasa base de la escala de Distimia del MCMI-II; BDI-II = Puntuación directa del Inventario de Depresión de Beck-II;  $z$  = Estadístico  $z$ ;  $p$  = Nivel de significación estadística.

**Puntos de corte del MCMI-III y del BDI-II para discriminar entre pacientes con un diagnóstico de trastorno depresivo mayor y pacientes con otro tipo de trastorno mental no depresivo según el DSM-IV**

Se recogen a continuación (véanse las Tablas 7.32, 7.33 y 7.34) las coordenadas de las curvas ROC de las escalas de CC y D del MCMI-III y de la curva ROC del BDI-II respecto al diagnóstico de trastorno depresivo mayor realizado por el clínico y reflejado en la ficha clínica (pacientes con otros trastornos depresivos excluidos). Estas coordenadas representan la sensibilidad y especificidad específica de cada una de las puntuaciones directas y puntuaciones tasa base (TB) de las escalas. En la Tabla 7.32 se puede observar que sensibilidades superiores a 0.70 para el diagnóstico de trastorno depresivo mayor se alcanzaban con puntuaciones directas y de tasa base en la escala de CC del MCMI-III inferiores a 9 y 63, respectivamente, mientras que especificidades superiores a 0.70 suponían puntuaciones directas y de tasa base superiores a 8 y 62, respectivamente.

**Tabla 7.32.** Coordenadas de las curvas ROC de las puntuaciones directas y de tasa base de la escala de Depresión Mayor del MCMI-III con respecto al diagnóstico de trastorno depresivo mayor (pacientes con otros trastornos depresivos excluidos)

Positivo si la puntuación directa de CC es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad	Positivo si la puntuación tasa base de CC MCMI-III es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad
-1.00	1.000	1.000	-1.00	1.000	1.000
0.50	1.000	0.878	2.50	1.000	0.911
1.50	0.976	0.772	6.50	1.000	0.894
2.50	0.976	0.691	8.50	1.000	0.846
3.50	0.976	0.634	9.50	1.000	0.829
4.50	0.952	0.585	11.00	0.976	0.829
5.50	0.881	0.504	12.50	0.976	0.789
6.50	0.881	0.423	14.00	0.976	0.780



Positivo si la puntuación directa de CC es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad	Positivo si la puntuación tasa base de CC MCMI-III es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad
7.50	0.833	0.341	17.00	0.976	0.707
8.50	0.738	0.285	19.50	0.976	0.699
9.50	0.643	0.228	21.50	0.976	0.691
10.50	0.571	0.163	23.50	0.976	0.642
11.50	0.476	0.122	27.00	0.976	0.634
12.50	0.452	0.114	33.00	0.976	0.610
13.50	0.381	0.073	37.00	0.976	0.602
14.50	0.286	0.065	41.50	0.905	0.545
15.50	0.214	0.065	46.50	0.881	0.488
16.50	0.190	0.024	50.00	0.881	0.463
17.50	0.143	0.024	52.50	0.881	0.455
18.50	0.071	0.016	54.50	0.833	0.398
19.50	0.071	0.008	57.00	0.833	0.390
20.50	0.048	0.008	58.50	0.833	0.382
21.50	0.024	0.008	59.50	0.833	0.374
50.00	0.000	0.008	61.00	0.786	0.301
79.00	0.000	0.000	62.50	0.714	0.276
			63.50	0.619	0.236
			64.50	0.595	0.236
			65.50	0.571	0.171
			66.50	0.548	0.171
			67.50	0.476	0.154
			68.50	0.476	0.114
			69.50	0.452	0.106
			70.50	0.429	0.089
			72.00	0.405	0.089
			73.50	0.333	0.049
			74.50	0.310	0.041
			76.00	0.214	0.041
			77.50	0.190	0.033
			79.00	0.190	0.016
			80.50	0.190	0.008
			82.00	0.143	0.000
			88.50	0.119	0.000
			94.50	0.071	0.000
			95.50	0.048	0.000
			104.00	0.024	0.000
			113.00	0.000	0.000

*Nota.* <sup>a</sup>El punto de corte menor es la mínima puntuación observada menos 1, mientras que el punto de corte mayor es la máxima puntuación observada más 1. Todos los demás puntos de corte son la media de dos puntuaciones observadas ordenadas y consecutivas.

En la Tabla 7.33 se puede observar que sensibilidades superiores a 0.70 para el diagnóstico del trastorno depresivo mayor se alcanzaban con puntuaciones directas y de tasa base en la escala de D del MCMI-III inferiores a 9 y 69, respectivamente, y puntuaciones directas y de tasa base superiores a 9 y 69, respectivamente, suponían especificidades superiores a 0.70 para dicho diagnóstico.

**Tabla 7.33.** Coordenadas de las curvas ROC de las puntuaciones directas y tasa base de la escala de Distimia del MCMI-III con respecto al diagnóstico de trastorno depresivo mayor (pacientes con otros trastornos depresivos excluidos)

Positivo si la puntuación directa de D es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad	Positivo si la puntuación tasa base de D MCMI-III es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad
-1.00	1.000	1.000	-1.00	1.000	1.000
0.50	0.976	0.886	1.00	1.000	0.935
1.50	0.976	0.829	2.50	0.976	0.935
2.50	0.976	0.748	4.00	0.976	0.927
3.50	0.952	0.683	6.00	0.976	0.911
4.50	0.952	0.634	7.50	0.976	0.902
5.50	0.952	0.545	8.50	0.976	0.846
6.50	0.929	0.480	9.50	0.976	0.837
7.50	0.833	0.439	12.50	0.976	0.821
8.50	0.762	0.358	15.50	0.976	0.780
9.50	0.643	0.276	18.00	0.976	0.764
10.50	0.619	0.220	21.50	0.976	0.740
11.50	0.548	0.187	25.00	0.976	0.724
12.50	0.476	0.146	28.50	0.952	0.724
13.50	0.429	0.122	34.00	0.952	0.642
14.50	0.333	0.098	39.00	0.952	0.577
15.50	0.310	0.065	42.50	0.952	0.561
16.50	0.262	0.049	47.50	0.929	0.528
17.50	0.143	0.041	51.00	0.929	0.520
18.50	0.048	0.008	52.50	0.929	0.512
19.50	0.024	0.000	55.00	0.881	0.472
21.00	0.000	0.000	57.50	0.857	0.463
			58.50	0.833	0.463
			59.50	0.833	0.455
			62.00	0.762	0.366
			64.50	0.762	0.358

Positivo si la puntuación directa de D es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad	Positivo si la puntuación tasa base de D MCMI-III es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad
			65.50	0.762	0.350
			67.00	0.738	0.333
			69.00	0.667	0.252
			70.50	0.667	0.244
			71.50	0.595	0.236
			72.50	0.595	0.220
			73.50	0.595	0.211
			74.50	0.548	0.211
			75.50	0.524	0.179
			76.50	0.452	0.171
			77.50	0.381	0.146
			78.50	0.357	0.114
			79.50	0.357	0.098
			80.50	0.310	0.073
			81.50	0.286	0.057
			83.50	0.238	0.057
			87.00	0.190	0.049
			89.50	0.167	0.049
			90.50	0.143	0.049
			92.00	0.143	0.041
			93.50	0.119	0.033
			94.50	0.119	0.024
			97.00	0.095	0.016
			99.50	0.095	0.008
			102.50	0.048	0.000
			105.50	0.024	0.000
			107.00	0.000	0.000

*Nota.* <sup>a</sup>El punto de corte menor es la mínima puntuación observada menos 1, mientras que el punto de corte mayor es la máxima puntuación observada más 1. Todos los demás puntos de corte son la media de dos puntuaciones observadas ordenadas y consecutivas.

En la Tabla 7.34 se puede observar que, con respecto a las puntuaciones del BDI-II, sensibilidades superiores a 0.70 para el diagnóstico del trastorno depresivo mayor se alcanzaban con puntuaciones inferiores a 23, mientras que especificidades superiores a 0.70 suponían puntuaciones superiores a 23.

**Tabla 7.34.** Coordenadas de las curvas ROC de las puntuaciones del BDI-II con respecto al diagnóstico de trastorno depresivo mayor (pacientes con otros trastornos depresivos excluidos)

Positivo si la puntuación total del BDI-II es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad
0.00	1.000	1.000
2.00	1.000	0.983
3.50	1.000	0.966
4.50	1.000	0.958
5.50	1.000	0.924
6.50	1.000	0.874
7.50	1.000	0.857
8.50	1.000	0.832
9.50	0.975	0.807
10.50	0.950	0.790
11.50	0.950	0.739
12.50	0.950	0.689
13.50	0.950	0.622
14.50	0.950	0.588
15.50	0.950	0.580
16.50	0.925	0.555
17.50	0.900	0.513
18.50	0.875	0.437
19.50	0.825	0.403
20.50	0.800	0.370
21.50	0.750	0.361
22.50	0.725	0.311
23.50	0.625	0.269
24.50	0.625	0.193
25.50	0.575	0.193
26.50	0.525	0.168
27.50	0.450	0.160
28.50	0.450	0.143
29.50	0.425	0.126
30.50	0.400	0.109
31.50	0.400	0.092
32.50	0.350	0.092
33.50	0.325	0.076
35.00	0.300	0.076
37.50	0.275	0.067
40.00	0.275	0.059
41.50	0.275	0.034
42.50	0.225	0.034
43.50	0.200	0.025

<b>Positivo si la puntuación total del BDI-II es mayor o igual que<sup>a</sup></b>	<b>Sensibilidad</b>	<b>1 - especificidad</b>
44.50	0.125	0.025
45.50	0.100	0.025
46.50	0.075	0.025
48.00	0.050	0.025
50.00	0.000	0.017
52.00	0.000	0.000

*Nota.* <sup>a</sup>El punto de corte menor es la mínima puntuación observada menos 1, mientras que el punto de corte mayor es la máxima puntuación observada más 1. Todos los demás puntos de corte son la media de dos puntuaciones observadas ordenadas y consecutivas.

En las Tablas 7.35, 7.36 y 7.37 se recogen más índices de validez diagnóstica en relación con el diagnóstico del trastorno depresivo mayor frente a otro trastorno mental no depresivo realizado por el clínico y reflejado en la ficha clínica o diagnóstica, para aquellas puntuaciones de las escalas de CC y D del MCMI-III y del BDI-II que han sido propuestas en los manuales correspondientes como puntos de corte para el diagnóstico de trastorno depresivo mayor (TB 75 y TB 85 del MCMI-III) o para la delimitación de diferentes niveles de gravedad de la sintomatología depresiva (BDI-II = 14, 20 y 28). Así mismo, se presentan esos mismos índices de validez diagnóstica para algunas puntuaciones que se encuentran cercanas a las anteriores.

El punto de corte TB 75 de la escala de CC del MCMI-III (véase Tabla 7.35) no fue el que tenía los mayores índices de concordancia o de Youden (0.33 y 26.88 respectivamente). Este punto de corte tenía una alta especificidad (95.93%), pero una sensibilidad baja (30.95%), aunque también fue la puntuación con mayor eficacia diagnóstica (79.39). El punto de corte TB 85 tuvo un índice de concordancia pobre (0.17), con un 100% de especificidad (todos los sujetos sin diagnóstico de trastorno depresivo mayor puntuaban por debajo de TB 85) a costa de una muy baja sensibilidad (11.9%). Por tanto, utilizando cualquiera de estos dos puntos de corte quedarían sin detectar un alto porcentaje de personas que sí tendrían un trastorno depresivo mayor. Otras

puntuaciones como TB 61 y TB 62 mostraban un índice kappa más elevado, 0.40, en el límite inferior del rango de concordancia moderada (0.40-0.60; Altman, 1991), con un mayor equilibrio entre sensibilidad y especificidad, cercanos al 70% (78.57% y 69.11% respectivamente) y con el índice de Youden más elevado (48.49, superior al mínimo recomendado para un instrumento diagnóstico). Para estos puntos de corte, el poder predictivo positivo no llegó a superar el 50% (47.14%), pero tuvieron un poder predictivo negativo muy alto (90.53%).

La puntuación directa 11 de la escala de CC tuvo valores de concordancia similares a TB 62 (kappa = 0.40). En este punto de corte se encontró una mayor especificidad (83.74%) que sensibilidad (57.14%), pero su índice de Youden de 40.88 fue adecuado; mientras que en el punto de corte de la puntuación directa 7, se encontró una mayor sensibilidad (88.1%) que especificidad (57.72%), con el índice de Youden más alto (45.82), pero con un kappa débil (0.33).

En el caso de la escala D (véase la Tabla 7.36), las puntuaciones TB 75 y TB 85, también tuvieron índices kappa débiles (0.32 y 0.23 respectivamente). Mientras que el punto de corte TB 75 tuvo un mayor equilibrio entre sensibilidad y especificidad (54.76% y 78.86% respectivamente), la puntuación TB 85 tuvo un mayor especificidad (94.31%) a costa de una débil sensibilidad (23.81%). Por tanto, se encontró de nuevo que las puntuaciones TB 75 y 85 dejaban sin detectar un alto porcentaje de pacientes que sí tenían el trastorno depresivo mayor. La puntuación TB 66 fue la puntuación TB de la escala D del MCMI-III que tuvo un mayor índice kappa (0.33) e índice de Youden (41.23), con una sensibilidad de 76.19% y una especificidad de 65.04%, mientras que la puntuación directa 11 mostró un índice kappa superior, aunque débil (0.37), con una especificidad de 78.05% y una sensibilidad de 61.9%, con el segundo índice de Youden más elevado (39.95).

**Tabla 7.35.** Índices de validez diagnóstica de distintas puntuaciones directas (PD) o tasa base (TB) de la escala de Depresión Mayor (CC) del MCMI-III para el diagnóstico de trastorno depresivo mayor (excluidos pacientes con otros trastornos depresivos)

Punto de corte	Sensibilidad	Especificidad	Índice de Youden	Poder predictivo positivo	Poder predictivo negativo	Tasa de falsos positivos	Tasa de falsos negativos	Kappa	Eficacia
<b>TB 75</b>	<b>30.95</b>	<b>95.93</b>	<b>26.88</b>	<b>72.22</b>	<b>80.27</b>	<b>4.07</b>	<b>69.05</b>	<b>0.33</b>	<b>79.39</b>
<b>TB 85</b>	<b>11.9</b>	<b>100</b>	<b>11.9</b>	<b>100</b>	<b>76.88</b>	<b>0</b>	<b>88.1</b>	<b>0.17</b>	<b>77.58</b>
TB 66	57.14	82.93	40.07	53.33	85	17.07	42.86	0.39	76.36
TB 67	54.76	82.93	37.69	52.27	84.3	17.07	45.24	0.37	75.76
TB 68	47.62	84.55	32.17	51.28	82.54	15.45	52.38	0.33	75.15
<b>TB 62</b>	<b>78.57</b>	<b>69.92</b>	<b>48.49</b>	<b>47.14</b>	<b>90.53</b>	<b>30.08</b>	<b>21.43</b>	<b>0.4</b>	<b>72.12</b>
<b>TB 61</b>	<b>78.57</b>	<b>69.92</b>	<b>48.49</b>	<b>47.14</b>	<b>90.53</b>	<b>30.08</b>	<b>21.43</b>	<b>0.4</b>	<b>72.12</b>
TB 60	83.33	62.6	45.93	43.21	91.67	37.4	16.67	0.35	67.88
<b>PD 7</b>	<b>88.1</b>	<b>57.72</b>	<b>45.82</b>	<b>41.57</b>	<b>93.42</b>	<b>42.28</b>	<b>11.9</b>	<b>0.33</b>	<b>65.45</b>
PD 9	73.81	71.54	45.35	46.97	88.89	28.46	26.19	0.38	72.12
PD 8	83.33	65.85	49.18	45.45	92.05	34.15	16.67	0.39	70.3
PD 10	64.29	77.24	41.53	49.09	86.36	22.76	35.71	0.38	73.94
<b>PD 11</b>	<b>57.14</b>	<b>83.74</b>	<b>40.88</b>	<b>54.55</b>	<b>85.12</b>	<b>16.26</b>	<b>42.86</b>	<b>0.4</b>	<b>76.97</b>
PD 12	47.62	87.8	35.42	57.14	83.08	12.2	52.38	0.37	77.58
PD 13	45.24	88.62	33.86	57.58	82.58	11.38	54.76	0.36	77.58
PD 14	38.1	92.68	30.78	64	81.43	7.32	61.9	0.36	78.79
PD 15	28.57	93.5	22.07	60	79.31	6.5	71.43	0.27	76.97
PD 16	21.43	93.5	14.93	52.94	77.7	6.5	78.57	0.19	75.15
PD 17	19.05	97.56	16.61	72.73	77.92	2.44	80.95	0.22	77.58
PD 18	14.29	97.56	11.85	66.67	76.92	2.44	85.71	0.16	76.36
PD 19	7.14	98.37	5.51	60	75.63	1.63	92.86	0.08	75.15
PD 20	7.14	99.19	6.33	75	75.78	0.81	92.86	0.09	75.76
PD 23	0	99.19	-0.81	0	74.39	0.81	100	-0.01	73.94
PD 28	0	99.19	-0.81	0	74.39	0.81	100	-0.01	73.94
PD 32	0	99.19	-0.81	0	74.39	0.81	100	-0.01	73.94
PD 35	0	99.19	-0.81	0	74.39	0.81	100	-0.01	73.94

**Tabla 7.36.** Índices de validez diagnóstica de distintas puntuaciones directas (PD) o tasa base (TB) de la escala de Distimia (D) del MCMI-III para el diagnóstico del trastorno depresivo mayor (excluidos pacientes con otros trastornos depresivos)

Punto de corte	Sensibilidad	Especificidad	Índice de Youden	Poder predictivo positivo	Poder predictivo negativo	Tasa de falsos positivos	Tasa de falsos negativos	Kappa	Eficacia
<b>TB 85</b>	<b>23.81</b>	<b>94.31</b>	<b>18.12</b>	<b>58.82</b>	<b>78.38</b>	<b>5.69</b>	<b>76.19</b>	<b>0.23</b>	<b>76.36</b>
<b>TB 75</b>	<b>54.76</b>	<b>78.86</b>	<b>33.62</b>	<b>46.94</b>	<b>83.62</b>	<b>21.14</b>	<b>45.24</b>	<b>0.32</b>	<b>72.73</b>
<b>TB 66</b>	<b>76.19</b>	<b>65.04</b>	<b>41.23</b>	<b>42.67</b>	<b>88.89</b>	<b>34.96</b>	<b>23.81</b>	<b>0.33</b>	<b>67.88</b>
TB 67	73.81	66.67	40.48	43.06	88.17	33.33	26.19	0.33	68.48
TB 68	73.81	66.67	40.48	43.06	88.17	33.33	26.19	0.33	68.48
TB 60	83.33	54.47	37.8	38.46	90.54	45.53	16.67	0.27	61.82
TB 61	76.19	63.41	39.6	41.56	88.64	36.59	23.81	0.31	66.67
TB 62	76.19	63.41	39.6	41.56	88.64	36.59	23.81	0.31	66.67
PD 8	83.33	56.1	39.43	39.33	90.79	43.9	16.67	0.29	63.03
PD 9	76.19	64.23	40.42	42.11	88.76	35.77	23.81	0.32	67.27
PD 10	64.29	72.36	36.65	44.26	85.58	27.64	35.71	0.32	70.3
<b>PD 11</b>	<b>61.9</b>	<b>78.05</b>	<b>39.95</b>	<b>49.06</b>	<b>85.71</b>	<b>21.95</b>	<b>38.1</b>	<b>0.37</b>	<b>73.94</b>
PD 15	33.33	90.24	23.57	53.85	79.86	9.76	66.67	0.27	75.76
PD 18	14.29	95.93	10.22	54.55	76.62	4.07	85.71	0.14	75.15
PD 20	2.38	100	2.38	100	75	0	97.62	0.04	75.15
PD 23	0	100	0	100	74.55	0	100	0	74.55
PD 28	0	100	0	100	74.55	0	100	0	74.55
PD 32	0	100	0	100	74.55	0	100	0	74.55
PD 35	0	100	0	100	74.55	0	100	0	74.55



**Tabla 7.37.** Índices de validez diagnóstica de distintas puntuaciones del BDI-II con respecto al diagnóstico de trastorno depresivo mayor (excluidos pacientes con el diagnóstico de otros trastornos depresivos)

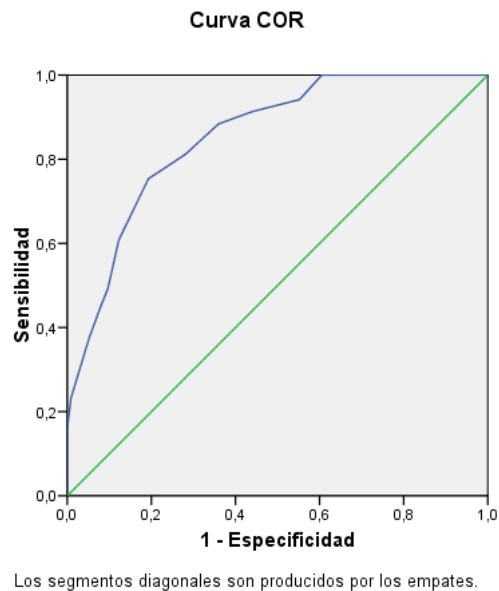
Punto de corte	Sensibilidad	Especificidad	Índice de Youden	Poder predictivo positivo	Poder predictivo negativo	Tasa de falsos positivos	Tasa de falsos negativos	Kappa	Eficacia
14	95	37.82	32.82	33.93	95.74	62.18	5	0.21	52.2
15	95	41.18	36.18	35.19	96.08	58.82	5	0.23	54.72
16	95	42.02	37.02	35.51	96.15	57.98	5	0.24	55.35
17	92.5	44.54	37.04	35.92	94.64	55.46	7.5	0.24	56.6
18	90	48.74	38.74	37.11	93.55	51.26	10	0.26	59.12
19	87.5	56.3	43.8	40.23	93.06	43.7	12.5	0.32	64.15
20	82.5	59.66	42.16	40.74	91.03	40.34	17.5	0.31	65.41
<b>21</b>	<b>80</b>	<b>63.03</b>	<b>43.03</b>	<b>42.11</b>	<b>90.36</b>	<b>36.97</b>	<b>20</b>	<b>0.33</b>	<b>67.3</b>
22	75	64.71	39.71	41.67	88.51	35.29	25	0.31	67.3
23	72.5	68.91	41.41	43.94	88.17	31.09	27.5	0.34	69.81
24	62.5	73.11	35.61	43.86	85.29	26.89	37.5	0.31	70.44
<b>25</b>	<b>62.5</b>	<b>80.67</b>	<b>43.17</b>	<b>52.08</b>	<b>86.49</b>	<b>19.33</b>	<b>37.5</b>	<b>0.4</b>	<b>76.1</b>
27	52.5	83.19	35.69	51.22	83.9	16.81	47.5	0.35	75.47
28	45	84.03	29.03	48.65	81.97	15.97	55	0.3	74.21
29	45	85.71	30.71	51.43	82.26	14.29	55	0.32	75.47
30	42.5	87.39	29.89	53.13	81.89	12.61	57.5	0.32	76.1
31	40	89.08	29.08	55.17	81.54	10.92	60	0.32	76.73
<b>32</b>	<b>40</b>	<b>90.76</b>	<b>30.76</b>	<b>59.26</b>	<b>81.82</b>	<b>9.24</b>	<b>60</b>	<b>0.34</b>	<b>77.99</b>
33	35	90.76	25.76	56	80.6	9.24	65	0.29	76.73
34	32.5	92.44	24.94	59.09	80.29	7.56	67.5	0.29	77.36
35	30	92.44	22.44	57.14	79.71	7.56	70	0.27	76.73

Con respecto al BDI-II (véase la Tabla 7.37), los distintos puntos de corte estudiados tuvieron índices kappa débiles (20-40; Altman, 1991). La puntuación total 25 tuvo el mayor índice kappa (0.40), que indicaba una concordancia moderada. Este punto de corte mostró una buena especificidad (80.67%), pero con una sensibilidad < 70% (62.5%). Además, fue precisamente el punto de corte 25 el que mostró un mayor índice de Youden (40.17) y tanto su poder predictivo positivo (52.08%) como su poder predictivo negativo (86.49%) superaron el 50%. En cambio, se puede seleccionar el punto de corte 21, aunque con un kappa débil (0.33), si se quiere primar la sensibilidad (80%) por encima de la especificidad (63.03%), para que el instrumento cumpla una de sus propiedades originales, la de servir como instrumento de *screening*. En cambio, el punto de corte 32 tuvo una alta especificidad (90.76%) a costa de una baja sensibilidad (40%), con un índice kappa también débil (0.34).

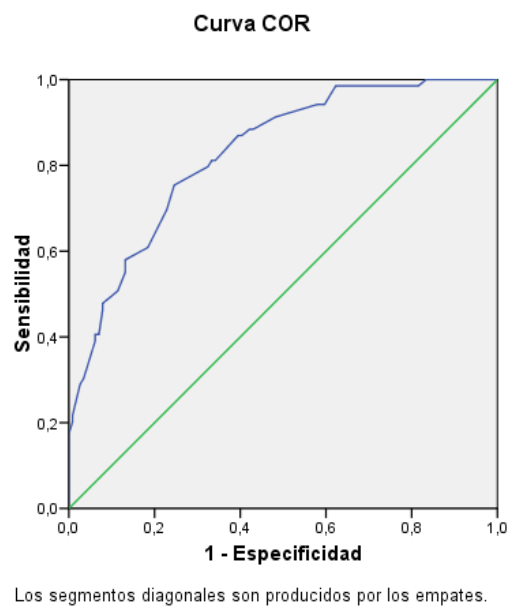
**Eficacia diagnóstica de las escalas de CC y D del MCMI-III y el BDI-II para discriminar entre pacientes con diagnóstico de trastorno depresivo mayor y pacientes con otro tipo de diagnóstico según el DSM-IV (diagnóstico según la entrevista SCID-I)**

Las curvas ROC realizadas sobre las puntuaciones directas y puntuaciones de TB de las escalas de CC y D del MCMI-III, así como sobre las puntuaciones totales del BDI-II, en relación con los diagnósticos de trastorno depresivo mayor realizados mediante la entrevista SCID-I, se recogen en las Figuras 7.19 a 7.23.

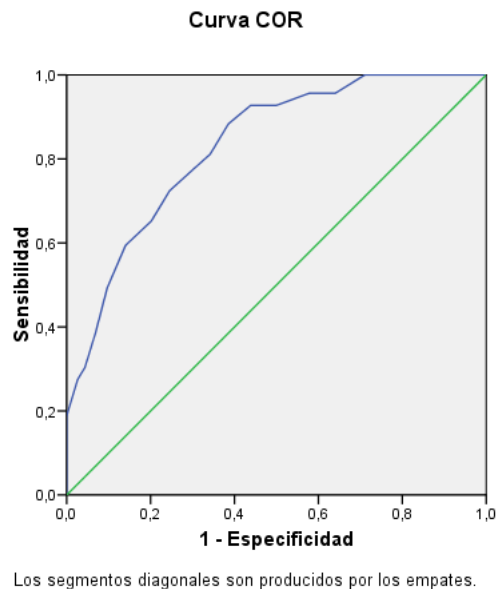
**Figura 7.19.** Curva ROC de las puntuaciones directas de la escala de CC del MCMI-III con respecto al diagnóstico de trastorno depresivo mayor (diagnóstico SCID-I)



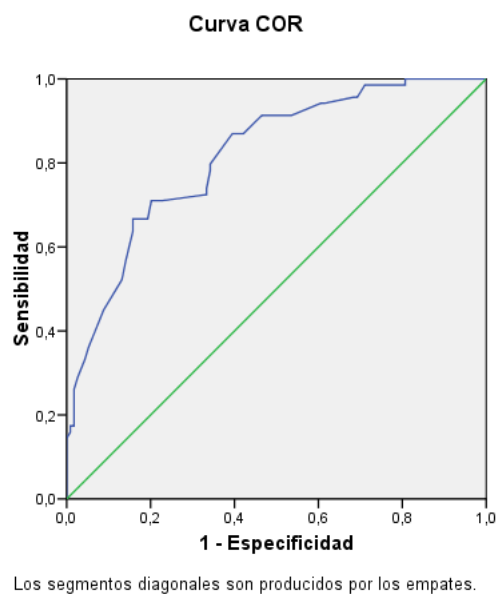
**Figura 7.20.** Curva ROC de las puntuaciones tasa base de la escala de CC del MCMI-III con respecto al diagnóstico de trastorno depresivo mayor (diagnóstico SCID-I)



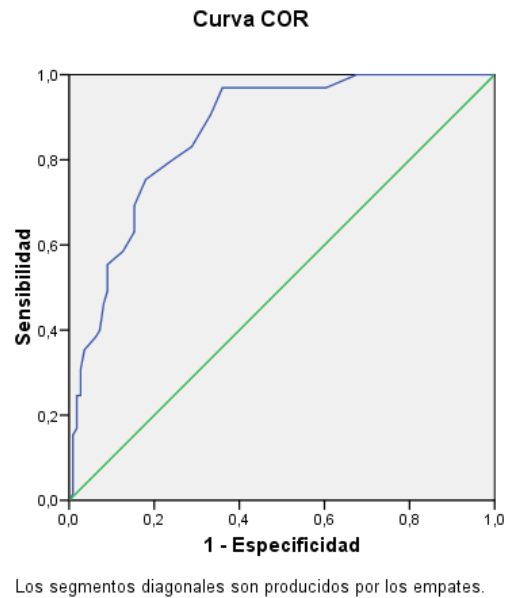
**Figura 7.21.** Curva ROC de las puntuaciones directas de la escala de D del MCMI-III con respecto al diagnóstico de trastorno depresivo mayor (diagnóstico SCID-I)



**Figura 7.22.** Curva ROC de las puntuaciones tasa base de la escala de D del MCMI-III con respecto al diagnóstico de Trastorno depresivo mayor (diagnóstico SCID-I)



**Figura 7.23.** Curva ROC de las puntuaciones del BDI-II con respecto al diagnóstico de trastorno depresivo mayor (diagnóstico SCID-I)



Se puede observar, a través de un análisis visual de las Figuras 7.19 a 7.23, que las áreas bajo la curva ROC de las escalas CC y D del MCMI-III se separaban moderadamente de los segmentos diagonales producidos por los empates entre sensibilidad y la tasa de falsos positivos (representados en las gráficas mediante la línea diagonal verde). De hecho, como se puede observar en la Tabla 7.38, el análisis estadístico indicaba que todas esas áreas se distanciaron de forma estadísticamente significativa del área de 0.50 que representa un diagnóstico basado en el azar y alcanzaron el criterio mínimo recomendado por algunos investigadores para un instrumento diagnóstico ( $> 0.70$ , según Swets, 1988), así como el criterio más estricto propuesto por otros investigadores ( $> 0.75$ , según Fan et al., 2006).

**Tabla 7.38.** Áreas bajo la curva ROC en relación con el diagnóstico de trastorno depresivo mayor (diagnóstico SCID-I) para las puntuaciones directas y tasa base de las escalas de Depresión Mayor y Distimia del MCMI-III y para las puntuaciones del BDI-II

Escala	Área ROC depresión mayor (SCID-I)	Intervalo de confianza asintótico del área al 95%	
		Límite inferior	Límite superior
<b>PD CC</b>	0.842*	0.784	0.899
<b>TB CC</b>	0.838*	0.781	0.895
<b>PD D</b>	0.825*	0.765	0.885
<b>TB D</b>	0.822*	0.761	0.883
<b>BDI-II</b>	0.860*	0.807	0.914

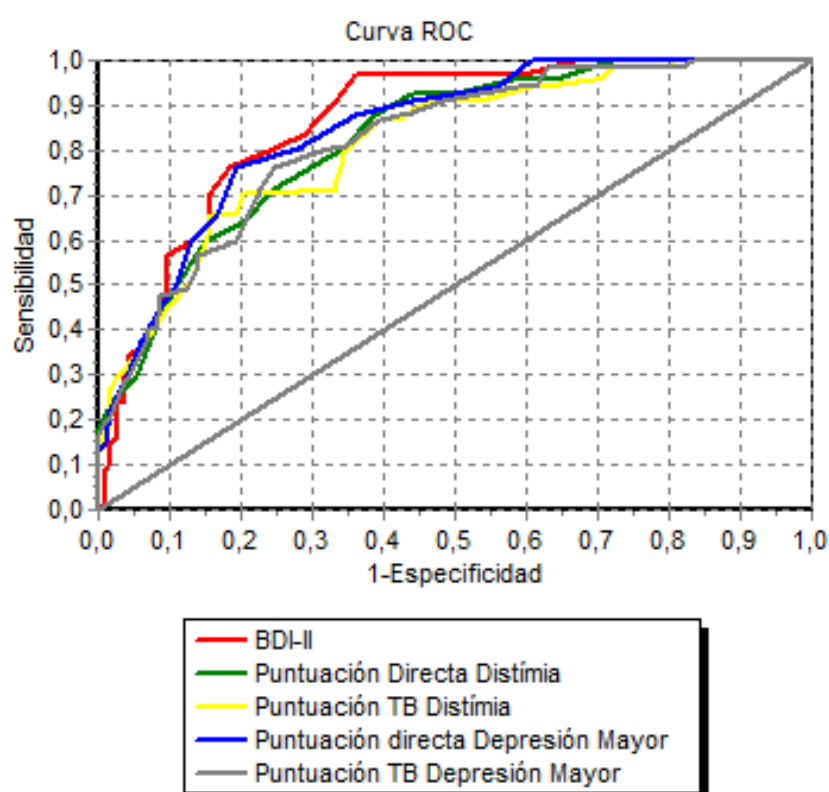
*Nota.* \* Estadísticamente significativas con  $p < 0,001$ . PD CC = Puntuación directa de la escala de Depresión Mayor del MCMI-III; TB CC = Tasa Base de la escala de Depresión Mayor del MCMI-III; PD D = Puntuación directa de la escala de Distimia del MCMI-III; TB D = Tasa Base de la escala de Distimia del MCMI-III BDI-II = Puntuación del Inventario de Depresión de Beck-II

El análisis visual de las Figuras 7.19 a 7.23 también parecía sugerir la impresión de que el área de la curva ROC del BDI-II se alejaba más de la línea que marca el diagnóstico realizado al azar que las áreas de las escalas de trastornos depresivos del MCMI-III, tanto de sus puntuaciones directas como de sus puntuaciones TB. Sin embargo, los datos recogidos en la Tabla 7.38 no confirmaban dicha impresión, puesto que los intervalos de confianza al 95% de las áreas bajo la curva ROC de los distintos instrumentos se solapaban.

De hecho, un análisis visual de comparación de las curvas ROC tampoco confirmaba esa impresión (véase la Figura 7.24) y, de manera más fehaciente, las pruebas  $z$  de comparación de pares de áreas bajo la curva ROC para el diagnóstico del

trastorno depresivo mayor según la SCID-I tampoco revelaron que las diferencias en dichas áreas entre las escalas de CC y D del MCMI-III o entre éstas y el BDI-II tuvieran significación estadística (véase la Tabla 7.39 y el Anexo 7.6).

**Figura 7.24.** Curvas ROC del BDI-II, las puntuaciones directas y TB de la escala de Distimia y Depresión Mayor del MCMI-III con respecto al diagnóstico de trastorno depresivo mayor (Diagnóstico SCID-I)



**Tabla 7.39.** Comparación de las áreas bajo la curva ROC en relación con el diagnóstico de trastorno depresivo mayor de la entrevista SCID-I para las puntuaciones directas y tasa base de las escalas de Depresión Mayor (CC) y Distimia (D) del MCMI-III y para las puntuaciones directas del BDI-II

Puntuación – Escala	TB – CC	PD – D	TB – D	BDI-II
<b>PD – CC</b>	$z = 0.287$ $p = 0.773$	$z = 0.669$ $p = 0.503$	$z = 0.717$ $p = 0.473$	$z = 0.571$ $p = 0.567$
<b>TB – CC</b>		$z = 0.520$ $p = 0.602$	$z = 0.594$ $p = 0.552$	$z = 0.640$ $p = 0.522$
<b>PD – D</b>			$z = 0.286$ $p = 0.775$	$z = 1.125$ $p = 0.260$
<b>TB – D</b>				$z = 0.669$ $p = 0.503$

*Nota.* PD – CC = Puntuación directa de la escala de Depresión Mayor del MCMI-II; TB – CC = Tasa base de la escala de Depresión Mayor del MCMI-II; PD – D = Puntuación directa de la escala de Distimia del MCMI-II; TB – D = Tasa base de la escala de Distimia del MCMI-II; BDI-II = Puntuación directa del Inventario de Depresión de Beck-II;  $z$  = Estadístico  $z$ ;  $p$  = Nivel de significación estadística.

**Puntos de corte del MCMI-III y del BDI-II para discriminar entre pacientes con un diagnóstico de trastorno depresivo mayor y pacientes con otro tipo de diagnóstico según el DSM-IV (diagnóstico según la entrevista SCID-I)**

Se recogen a continuación (véanse las Tablas 7.40, 7.41 y 7.42) las coordenadas de las curvas ROC de las escalas de CC y D del MCMI-III y del BDI-II respecto al diagnóstico de trastorno depresivo mayor realizado a través de la SCID-I. Estas coordenadas representan la sensibilidad y especificidad específicas de cada una de las puntuaciones (directas y tasa base) de las escalas de los trastornos depresivos del MCMI-III y del BDI-II.



En la Tabla 7.40 se muestran las coordenadas de las curvas ROC de las puntuaciones directas y tasa base de la escala de CC del MCMI-III respecto al diagnóstico de trastorno depresivo mayor de la SCID-I. Para la escala CC del MCMI-III, se alcanzaron sensibilidades superiores a 0.70 para el diagnóstico del trastorno depresivo mayor con puntuaciones directas y tasa base inferiores a 9 y 64, respectivamente, mientras que especificidades superiores a 0.70 suponían puntuaciones directas y de tasa base superiores a 7 y 61, respectivamente.

**Tabla 7.40.** Coordenadas de las curvas ROC de las puntuaciones directas y tasa base de la escala de Depresión Mayor del MCMI-III con respecto al diagnóstico de trastorno depresivo mayor de la SCID-I

Positivo si la puntuación directa de CC es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad	Positivo si la puntuación TB de CC MCMI-III es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad
-1.00	1.000	1.000	-1.00	1.000	1.000
0.50	1.000	0.878	2.50	1.000	0.913
1.50	1.000	0.748	6.50	1.000	0.896
2.50	1.000	0.670	8.50	1.000	0.843
3.50	1.000	0.609	9.50	1.000	0.826
4.50	0.942	0.557	11.00	1.000	0.817
5.50	0.913	0.443	12.50	1.000	0.765
6.50	0.884	0.365	14.00	1.000	0.757
7.50	0.812	0.287	17.00	1.000	0.687
8.50	0.754	0.200	19.50	1.000	0.678
9.50	0.667	0.165	21.50	1.000	0.670
10.50	0.594	0.130	23.50	1.000	0.626
11.50	0.478	0.104	27.00	1.000	0.617
12.50	0.435	0.087	33.00	0.957	0.591
13.50	0.362	0.061	37.00	0.957	0.574
14.50	0.275	0.035	41.50	0.928	0.478
15.50	0.217	0.017	45.50	0.899	0.426
16.50	0.145	0.009	47.00	0.899	0.417
17.50	0.116	0.009	50.00	0.884	0.400
18.50	0.058	0.009	52.50	0.884	0.391
19.50	0.043	0.009	54.50	0.826	0.339
20.50	0.029	0.009	57.50	0.826	0.330

Positivo si la puntuación directa de CC es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad	Positivo si la puntuación TB de CC MCMI-III es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad
21.50	0.014	0.009	59.50	0.812	0.322
50.00	0.000	0.009	61.00	0.768	0.243
79.00	0.000	0.000	62.50	0.710	0.226
			64.00	0.623	0.183
			65.50	0.594	0.130
			66.50	0.565	0.130
			67.50	0.507	0.113
			68.50	0.478	0.078
			69.50	0.464	0.078
			70.50	0.406	0.070
			71.50	0.406	0.061
			72.50	0.391	0.061
			73.50	0.304	0.035
			74.50	0.290	0.026
			75.50	0.217	0.009
			76.50	0.203	0.009
			77.50	0.174	0.000
			79.00	0.145	0.000
			80.50	0.130	0.000
			82.00	0.087	0.000
			88.50	0.072	0.000
			94.50	0.043	0.000
			95.50	0.029	0.000
			104.00	0.014	0.000
			113.00	0.000	0.000

*Nota.* <sup>a</sup>El punto de corte menor es la mínima puntuación observada menos 1, mientras que el punto de corte mayor es la máxima puntuación observada más 1. Todos los demás puntos de corte son la media de dos puntuaciones observadas ordenadas y consecutivas.

En la Tabla 7.41 se recogen las coordenadas de las curvas ROC de las puntuaciones directas y tasa base de la escala de D del MCMI-III respecto al diagnóstico de trastorno depresivo mayor según la SCID-I. Dichas coordenadas muestran que se alcanzaban sensibilidades superiores a 0.70 para el diagnóstico del trastorno depresivo mayor con puntuaciones directas y de tasa base inferiores a 10 y 71 en la escala D, respectivamente, mientras que especificidades superiores a 0.70 suponían puntuaciones directas y de tasa base en la escala D superiores a 9 y 69, respectivamente.

**Tabla 7.41.** Coordenadas de las curvas ROC de las puntuaciones directas y tasa base de la escala de Distimia del MCMI-III con respecto al diagnóstico de trastorno depresivo mayor de la SCID-I

Positivo si la puntuación directa de D MCMI-III es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad	Positivo si la puntuación TB de D MCMI-III es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad
-1.00	1.000	1.000	-1.00	1.000	1.000
0.50	1.000	0.878	1.00	1.000	0.939
1.50	1.000	0.817	2.50	1.000	0.930
2.50	1.000	0.713	4.00	1.000	0.922
3.50	0.957	0.643	6.00	1.000	0.904
4.50	0.957	0.583	7.50	1.000	0.896
5.50	0.928	0.496	8.50	1.000	0.835
6.50	0.928	0.435	9.50	1.000	0.826
7.50	0.884	0.383	12.50	1.000	0.809
8.50	0.812	0.339	15.50	1.000	0.757
9.50	0.725	0.243	18.00	1.000	0.739
10.50	0.652	0.200	21.50	1.000	0.704
11.50	0.594	0.139	25.00	0.971	0.687
12.50	0.478	0.096	28.50	0.971	0.678
13.50	0.377	0.070	30.50	0.957	0.609
14.50	0.290	0.043	34.50	0.957	0.600
15.50	0.261	0.026	39.00	0.928	0.530
16.50	0.232	0.017	42.50	0.928	0.504
17.50	0.174	0.000	47.50	0.928	0.470
18.50	0.058	0.000	51.50	0.928	0.461
19.50	0.014	0.000	55.00	0.884	0.417
21.00	0.000	0.000	57.50	0.884	0.409
			58.50	0.884	0.400
			59.50	0.884	0.391
			62.50	0.812	0.339
			65.50	0.797	0.339
			66.50	0.754	0.330
			67.50	0.739	0.330
			69.00	0.725	0.226
			70.50	0.725	0.200
			71.50	0.667	0.191
			72.50	0.667	0.165
			73.50	0.667	0.157
			74.50	0.638	0.157
			75.50	0.565	0.139

Positivo si la puntuación directa de D MCMI-III es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad	Positivo si la puntuación TB de D MCMI-III es mayor o igual que <sup>a</sup>	Sensibilidad	1 - especificidad
			76.50	0.522	0.130
			77.50	0.449	0.087
			78.50	0.362	0.052
			79.50	0.333	0.043
			80.50	0.290	0.026
			81.50	0.261	0.017
			83.50	0.217	0.017
			87.00	0.188	0.017
			89.50	0.174	0.017
			90.50	0.174	0.009
			92.00	0.159	0.009
			93.50	0.145	0.000
			94.50	0.130	0.000
			97.00	0.101	0.000
			99.50	0.087	0.000
			100.50	0.043	0.000
			103.00	0.029	0.000
			105.50	0.014	0.000
			107.00	0.000	0.000

*Nota.* <sup>a</sup>El punto de corte menor es la mínima puntuación observada menos 1, mientras que el punto de corte mayor es la máxima puntuación observada más 1. Todos los demás puntos de corte son la media de dos puntuaciones observadas ordenadas y consecutivas. CC MCMI-III = Escala de Depresión Mayor del MCMI-III. D MCMI-III = Escala de Distimia del MCMI-III.

Finalmente, en la Tabla 7.42 se recogen las coordenadas de la curva ROC de las puntuaciones del BDI-II con respecto al diagnóstico de trastorno depresivo mayor de la SCID-I, las cuales indican que sensibilidades superiores a 0.70 se alcanzaban con puntuaciones del BDI-II inferiores a 24, mientras que especificidades superiores a 0.70 suponían puntuaciones superiores a 21.

**Tabla 7.42.** Coordenadas de las curvas ROC de las puntuaciones del BDI-II con respecto al diagnóstico de trastorno depresivo mayor de la SCID-I

<b>Positivo si la puntuación total del BDI-II es mayor o igual que<sup>a</sup></b>	<b>Sensibilidad</b>	<b>1 - especificidad</b>
0.00	1.000	1.000
2.00	1.000	0.982
3.50	1.000	0.964
4.50	1.000	0.955
5.50	1.000	0.920
6.50	1.000	0.866
7.50	1.000	0.848
8.50	1.000	0.830
9.50	1.000	0.795
10.50	1.000	0.777
11.50	1.000	0.732
12.50	1.000	0.679
13.50	0.969	0.607
14.50	0.969	0.554
15.50	0.969	0.545
16.50	0.969	0.509
17.50	0.969	0.446
18.50	0.969	0.366
19.50	0.908	0.339
20.50	0.862	0.313
21.50	0.831	0.295
22.50	0.800	0.250
23.50	0.754	0.188
24.50	0.692	0.161
25.50	0.631	0.161
26.50	0.585	0.134
27.50	0.554	0.098
28.50	0.492	0.098
29.50	0.462	0.089
30.50	0.400	0.080
31.50	0.385	0.071
32.50	0.354	0.045
33.50	0.308	0.036
34.50	0.292	0.036
35.50	0.277	0.036
37.50	0.246	0.036
39.50	0.246	0.027
40.50	0.231	0.027
41.50	0.200	0.018
42.50	0.169	0.018
43.50	0.154	0.009

<b>Positivo si la puntuación total del BDI-II es mayor o igual que<sup>a</sup></b>	<b>Sensibilidad</b>	<b>1 - especificidad</b>
44.50	0.108	0.009
45.50	0.092	0.009
46.50	0.077	0.009
48.00	0.062	0.009
50.00	0.015	0.009
52.00	0.000	0.000

*Nota.* <sup>a</sup>El punto de corte menor es la mínima puntuación observada menos 1, mientras que el punto de corte mayor es la máxima puntuación observada más 1. Todos los demás puntos de corte son la media de dos puntuaciones observadas ordenadas y consecutivas.

En las Tablas 7.43, 7.44 y 7.45 se recogen más índices de validez diagnóstica en relación con los diagnósticos de trastorno depresivo mayor realizados mediante la entrevista SCID-I para aquellas puntuaciones de las escalas de CC y D del MCMI-III y del BDI-II que han sido propuestas en los manuales correspondientes como puntos de corte para el diagnóstico de trastorno depresivo mayor (TB 75 y TB 85 del MCMI-II) o para la delimitación de diferentes niveles de gravedad de la sintomatología depresiva (BDI-II = 14, 20 y 28). Así mismo, se presentan esos mismos índices de validez diagnóstica para algunas puntuaciones que se encuentran cercanas a las anteriores.

Se puede apreciar, en el caso de la escala de CC del MCMI-III, que las puntuaciones TB 75 y TB 85 no fueron las puntuaciones con mejores índices de eficacia diagnóstica (véase la Tabla 7.43). Estos puntos de corte tuvieron una débil sensibilidad (28.99% y 7.25%, respectivamente), lo que indicaba que existía un alto porcentaje de pacientes diagnosticados con trastorno depresivo mayor mediante la SCID-I que no habían sido detectados por la escala CC del MCMI-III. Dichos puntos de corte mostraban, en cambio, una alta especificidad (97.37% y 100% respectivamente) y poder predictivo positivo (86.96% y 100% respectivamente) y un moderado poder predictivo negativo

(69.38% y 64.04% respectivamente), poderes predictivos que eran superiores en ambos casos al 50%. Sin embargo, la puntuación TB 75 también mostraba un débil índice de concordancia (0.3), el cuál era muy pobre en el caso de la puntuación TB 85 (0.09) (Altman, 1991).

Otras puntuaciones TB en la escala de CC tuvieron índices kappa moderados (0.40-0.60; Altman, 1991), como las puntuaciones TB 61 y TB 62 (kappa = 0.51), con unos índices de Youden de 52.25 (superiores a 40, considerado como el valor estándar para un instrumento diagnóstico). También se encontró en estos puntos de corte un equilibrio entre sensibilidad y especificidad, superiores ambas a 70% (76.81% y 75.44%, respectivamente), y un poder predictivo negativo mayor que el poder predictivo positivo (84.31% y 65.43%, respectivamente), pero ambos superiores a 50%. Entre las puntuaciones directas de la escala CC, la puntuación 9 mostró el mayor índice kappa (0.55), con una alta especificidad (80.7%) y sensibilidad (75.36%) y con el mayor índice de Youden (56.6). En este punto de corte también el poder predictivo negativo (84.4%) fue mayor que el poder predictivo positivo (70.27%), pero ambos de nuevo se situaban por encima del 50%.

Por el contrario, en la escala de D del MCMI-III (véase la Tabla 7.44), el punto de corte con un mayor índice kappa fue la puntuación TB 75, con 0.48 (índice kappa moderado; Altman, 1991). En este punto de corte se encontró una alta especificidad (83.62%), pero una sensibilidad insuficiente (63.67%), por debajo ligeramente del valor de referencia de 70%. La puntuación TB 75 fue también el punto de corte de la escala D con el índice de Youden más elevado (47.39). La puntuación TB 85 de la escala D tuvo un valor ligeramente más elevado de especificidad (98.26%) a cambio de perder mucha sensibilidad (21.74%), de manera que el índice kappa de este punto fue débil (0.23). Otros puntos de corte alternativos con buen índice kappa como, por ejemplo, la

puntuación TB 60, mostró un índice kappa también moderado (0.45) y priorizaba la sensibilidad (88.41%) sobre la especificidad (60.87%).

La puntuación directa 10 de la escala de D del MCMI-III tuvo un índice kappa moderado (0.47) con respecto al diagnóstico de trastorno depresivo mayor de la entrevista SCID-I y la puntuación directa 8 tuvo el índice de Youden más elevado para la escala D (50.15), también con un índice kappa moderado (0.45) . Mientras que la puntuación directa 8 tuvo una sensibilidad más elevada que la especificidad (88.41% y 61.74% respectivamente), la puntuación directa 10 tuvo una sensibilidad y especificidad superiores al 70% y más equilibradas entre ellas (72.46% y 75.65%, respectivamente).

En el caso del BDI-II (como se puede ver en la Tabla 7.45), se encontró en la puntuación 25 el valor de kappa más elevado, 0.57, en el rango superior del índice de kappa moderado (40-60; Altaman, 1991). Este punto de corte mostró datos equilibrados de sensibilidad en relación con la especificidad (75.38% y 81.98%, respectivamente), ambos superiores al estándar mínimo para ser considerados buen instrumento diagnóstico. Además, el punto de corte 25 tuvo un aceptable índice de Youden (57.36) y el poder predictivo positivo y el poder predictivo negativo son notablemente superiores al 50% (71.01% y 85.05%, respectivamente). El punto de corte 19 del BDI-II tuvo también un kappa moderado de 0.54, con un índice de Youden aún mayor (60.88%), primando la sensibilidad (96.92%) por encima de la especificidad, menor esta última al 70% (63.96%).



**Tabla 7.43.** Índices de validez diagnóstica de distintas puntuaciones directas (PD) o tasa base (TB) de la escala de Depresión Mayor (CC) del MCMI-III con respecto al diagnóstico de trastorno depresivo mayor (diagnóstico SCID-I)

Punto de corte	Sensibilidad	Especificidad	Índice de Youden	Poder predictivo positivo	Poder predictivo negativo	Tasa de falsos positivos	Tasa de falsos negativos	Kappa	Eficacia
<b>TB 75</b>	<b>28.99</b>	<b>97.37</b>	<b>26.36</b>	<b>86.96</b>	<b>69.38</b>	<b>2.63</b>	<b>71.01</b>	<b>0.3</b>	<b>71.58</b>
<b>TB 85</b>	<b>7.25</b>	<b>100</b>	<b>7.25</b>	<b>100</b>	<b>64.04</b>	<b>0</b>	<b>92.75</b>	<b>0.09</b>	<b>65.03</b>
TB 66	59.42	86.84	46.26	73.21	77.95	13.16	40.58	0.48	76.5
TB 67	56.52	86.84	43.36	72.22	76.74	13.16	43.48	0.45	75.41
TB 68	50.72	88.6	39.32	72.92	74.81	11.4	49.28	0.42	74.32
<b>TB 62</b>	<b>76.81</b>	<b>75.44</b>	<b>52.25</b>	<b>65.43</b>	<b>84.31</b>	<b>24.56</b>	<b>23.19</b>	<b>0.51</b>	<b>75.96</b>
<b>TB 61</b>	<b>76.81</b>	<b>75.44</b>	<b>52.25</b>	<b>65.43</b>	<b>84.31</b>	<b>24.56</b>	<b>23.19</b>	<b>0.51</b>	<b>75.96</b>
TB 60	81.16	67.54	48.7	60.22	85.56	32.46	18.84	0.46	72.68
PD 7	88.41	64.04	52.45	59.8	90.12	35.96	11.59	0.48	73.22
<b>PD 9</b>	<b>75.36</b>	<b>80.7</b>	<b>56.06</b>	<b>70.27</b>	<b>84.4</b>	<b>19.3</b>	<b>24.64</b>	<b>0.55</b>	<b>78.69</b>
PD 8	81.16	71.93	53.09	63.64	86.32	28.07	18.84	0.5	75.41
PD 10	63.89	84.21	48.1	71.88	78.69	15.79	36.11	0.49	76.34
PD 11	59.42	87.72	47.14	74.55	78.13	12.28	40.58	0.49	77.05
PD 12	47.83	90.35	38.18	75	74.1	9.65	52.17	0.41	74.32
PD 13	43.48	92.11	35.59	76.92	72.92	7.89	56.52	0.39	73.77
PD 14	36.23	94.74	30.97	80.65	71.05	5.26	63.77	0.35	72.68
PD 15	27.54	97.37	24.91	86.36	68.94	2.63	72.46	0.29	71.04
PD 16	21.74	99.12	20.86	93.75	67.66	0.88	78.26	0.25	69.95
PD 17	14.49	100	14.49	100	65.9	0	85.51	0.17	67.76
PD 18	11.59	100	11.59	100	65.14	0	88.41	0.14	66.67
PD 19	5.8	100	5.8	100	63.69	0	94.2	0.07	64.48
PD 20	4.35	100	4.35	100	63.33	0	95.65	0.05	63.93
PD 23	0	100	0	100	62.3	0	100	0	62.3
PD 28	0	100	0	100	62.3	0	100	0	62.3
PD 32	0	100	0	100	62.3	0	100	0	62.3
PD 35	0	100	0	100	62.3	0	100	0	62.3

**Tabla 7.44.** Índices de validez diagnóstica de distintas puntuaciones directas (PD) o tasa base (TB) de la escala de Distimia (D) del MCMI-III con respecto al diagnóstico de trastorno depresivo mayor (diagnóstico SCID-I)

Punto de corte	Sensibilidad	Especificidad	Índice de Youden	Poder predictivo positivo	Poder predictivo negativo	Tasa de falsos positivos	Tasa de falsos negativos	Kappa	Eficacia
<b>TB 85</b>	<b>21.74</b>	<b>98.26</b>	<b>20</b>	<b>88.24</b>	<b>67.66</b>	<b>1.74</b>	<b>78.26</b>	<b>0.24</b>	<b>69.57</b>
<b>TB 75</b>	<b>63.77</b>	<b>83.62</b>	<b>47.39</b>	<b>69.84</b>	<b>79.51</b>	<b>16.38</b>	<b>36.23</b>	<b>0.48</b>	<b>76.22</b>
TB 66	79.71	66.09	45.8	58.51	84.44	33.91	20.29	0.43	71.2
TB 67	75.36	66.96	42.32	57.78	81.91	33.04	24.64	0.4	70.11
TB 68	73.91	66.96	40.87	57.3	81.05	33.04	26.09	0.39	69.57
<b>TB 60</b>	<b>88.41</b>	<b>60.87</b>	<b>49.28</b>	<b>57.55</b>	<b>89.74</b>	<b>39.13</b>	<b>11.59</b>	<b>0.45</b>	<b>71.2</b>
TB 61	81.16	66.09	47.25	58.95	85.39	33.91	18.84	0.44	71.74
TB 62	81.16	66.09	47.25	58.95	85.39	33.91	18.84	0.44	71.74
<b>PD 8</b>	<b>88.41</b>	<b>61.74</b>	<b>50.15</b>	<b>58.1</b>	<b>89.87</b>	<b>38.26</b>	<b>11.59</b>	<b>0.45</b>	<b>71.74</b>
PD 9	81.16	66.09	47.25	58.95	85.39	33.91	18.84	0.44	71.74
<b>PD 10</b>	<b>72.46</b>	<b>75.65</b>	<b>48.11</b>	<b>64.1</b>	<b>82.08</b>	<b>24.35</b>	<b>27.54</b>	<b>0.47</b>	<b>74.46</b>
PD 11	65.22	80	45.22	66.18	79.31	20	34.78	0.45	74.46
PD 15	28.99	95.65	24.64	80	69.18	4.35	71.01	0.28	70.65
PD 18	17.39	100	17.39	100	66.86	0	82.61	0.21	69.02
PD 20	1.45	100	1.45	100	62.84	0	98.55	0.02	63.04
PD 23	0	100	0	100	62.5	0	100	0	62.5
PD 28	0	100	0	100	62.5	0	100	0	62.5
PD 32	0	100	0	100	62.5	0	100	0	62.5
PD 35	0	100	0	100	62.5	0	100	0	62.5

**Tabla 7.45.** Índices de validez diagnóstica de distintas puntuaciones del BDI-II con respecto al diagnóstico de trastorno depresivo mayor (diagnóstico SCID-I)

Punto de corte	Sensibilidad	Especificidad	Índice de Youden	Poder predictivo positivo	Poder predictivo negativo	Tasa de falsos positivos	Tasa de falsos negativos	Kappa	Eficacia
14	96.92	39.64	36.56	48.46	95.65	60.36	3.08	0.3	60.8
15	96.92	45.05	41.97	50.81	96.15	54.95	3.08	0.35	64.2
16	96.92	45.95	42.87	51.22	96.23	54.05	3.08	0.36	64.77
17	96.92	49.55	46.47	52.94	96.49	50.45	3.08	0.4	67.05
18	96.92	55.86	52.78	56.25	96.88	44.14	3.08	0.46	71.02
<b>19</b>	<b>96.92</b>	<b>63.96</b>	<b>60.88</b>	<b>61.17</b>	<b>97.26</b>	<b>36.04</b>	<b>3.08</b>	<b>0.54</b>	<b>76.14</b>
20	90.77	66.67	57.44	61.46	92.5	33.33	9.23	0.52	75.57
21	86.15	69.37	55.52	62.22	89.53	30.63	13.85	0.51	75.57
22	83.08	71.17	54.25	62.79	87.78	28.83	16.92	0.51	75.57
23	80	75.68	55.68	65.82	86.6	24.32	20	0.53	77.27
<b>24</b>	<b>75.38</b>	<b>81.98</b>	<b>57.36</b>	<b>71.01</b>	<b>85.05</b>	<b>18.02</b>	<b>24.62</b>	<b>0.57</b>	<b>79.55</b>
25	69.23	84.68	53.91	72.58	82.46	15.32	30.77	0.54	78.98
27	58.46	87.39	45.85	73.08	78.23	12.61	41.54	0.48	76.7
28	55.38	90.99	46.37	78.26	77.69	9.01	44.62	0.49	77.84
29	49.23	90.99	40.22	76.19	75.37	9.01	50.77	0.43	75.57
30	46.15	91.89	38.04	76.92	74.45	8.11	53.85	0.41	75
31	40	92.79	32.79	76.47	72.54	7.21	60	0.36	73.3
32	38.46	93.69	32.15	78.13	72.22	6.31	61.54	0.36	73.3
33	35.38	96.4	31.78	85.19	71.81	3.6	64.62	0.36	73.86
34	30.77	97.3	28.07	86.96	70.59	2.7	69.23	0.32	72.73
35	29.23	97.3	26.53	86.36	70.13	2.7	70.77	0.31	72.16

#### 7.4. Discusión

Este estudio se llevó a cabo con la finalidad de conocer las propiedades diagnósticas de las escalas de Depresión Mayor y Distimia del MCMI-III. El manual de este inventario señala que el MCMI-III tiene varios niveles de interpretación y en un primer nivel trata de proporcionar información para hacer diagnósticos (Millon et al., 2007), pero, sin embargo, no hay estudios publicados sobre la eficacia diagnóstica de las escalas de la adaptación española del MCMI-III (Millon et al., 2007, adaptación española de V. Cardenal y M. P. Sánchez). Además, los estudios de eficacia diagnóstica realizados por los propios autores del MCMI-III (Millon, 1994; Millon, Davis y Millon, 1997) han generado una gran controversia (Hsu, 2002; Rezlaff, 1996) y tan solo se han publicado tres estudios independientes sobre la eficacia diagnóstica de sus escalas de trastornos depresivos, dos con la versión original del inventario (Gibeau y Choca, 2005; Saulusman, 2011) y uno con la adaptación belga para población flamencoparlante (Rossi y Sloore, 2005).

Por tanto, en este estudio prospectivo se pretendía, en primer lugar, examinar la validez diagnóstica de las puntuaciones TB 75 y TB 85 (las puntuaciones de corte estándar o habituales en el MCMI-III) de las escalas CC y D para el diagnóstico del trastorno depresivo mayor y para el diagnóstico del conjunto de trastornos depresivos. En segundo lugar, el estudio pretendía conocer si existían otras puntuaciones de las escalas CC y D del MCMI-III que tuvieran mejor validez diagnóstica que las puntuaciones de corte habituales de dichas escalas. Así mismo, el estudio también intentaba examinar si los índices de validez diagnóstica de las escalas CC y D del MCMI-III eran mejores que los de otros instrumentos para evaluar la depresión ampliamente utilizados, en concreto, mejores que los del BDI-II.

**¿Cuál es la validez diagnóstica de las puntuaciones TB 75 y TB 85 de las escalas CC y D del MCMI-III para el diagnóstico del trastorno depresivo mayor?**

Los resultados del presente estudio indican que, para el diagnóstico del trastorno depresivo mayor, la validez diagnóstica de las puntuaciones TB 75 y TB 85 de las escalas CC y D de la adaptación española del MCMI-III es, en general, cuestionable. Para mayor claridad, se han resumido esos resultados en la Tabla 7.46 y de los mismos se puede concluir que, en general: (1) los índices de concordancia diagnóstica (kappa) de las puntuaciones TB 75 y TB 85 de las escalas CC y D se situaban entre débiles ( $< 0,20$ ) y pobres ( $< 0,40$ ); (2) los valores de sensibilidad de dichas puntuaciones fueron significativamente inferiores al estándar considerado como aceptable ( $< 70\%$ ), y (3) los índices de Youden no reflejaban un equilibrio adecuado entre sensibilidad y especificidad ( $< 40$ ).

Bien es cierto que dichas puntuaciones obtuvieron valores de especificidad elevados, superiores a 70%, valores predictivos positivos más altos, en general, que 50%, y valores predictivos negativos superiores a 60% e incluso a 80%. Por tanto, según estos datos, especialmente los datos de especificidad y valor predictivo negativo, los puntos de corte TB 75 y TB 85 de las escalas de CC y D de la adaptación española del MCMI-III serían útiles para detectar correctamente a los pacientes que no tienen un diagnóstico de trastorno depresivo mayor. Es decir, un alto porcentaje de pacientes que efectivamente no tienen un trastorno depresivo mayor puntúan por debajo de esos puntos de corte y, complementariamente, si un paciente ha puntuado por debajo de esas puntuaciones TB, la probabilidad de que efectivamente no tenga un trastorno depresivo mayor es elevada. Sin embargo, los valores de sensibilidad de las puntuaciones TB 75 y TB 85 de las escalas de trastornos depresivos de la adaptación española del MCMI-III son inferiores a los valores aceptables para un instrumento diagnóstico e indican que

muchos pacientes que tienen un trastorno depresivo mayor puntúan por debajo de dichas puntuaciones y no son detectados por el instrumento.

Este patrón de resultados cuestionables sobre el rendimiento diagnóstico de las puntuaciones de corte habituales de las escalas CC y D del MCMI-III es evidente tanto si se considera la habilidad del MCMI-III para discriminar entre trastorno depresivo mayor y otros trastornos mentales como si se considera su habilidad para discriminar entre trastorno depresivo mayor y otros trastornos mentales no depresivos. Efectivamente, incluso siguiendo las recomendaciones de Ganellen (1996) de evaluar la eficacia diagnóstica del MCMI-III considerando su capacidad para discriminar un trastorno depresivo mayor frente a otro trastorno mental no depresivo, es decir, excluyendo de la muestra de pacientes del presente estudio los pacientes que tenían otros diagnósticos de depresión distintos al trastorno depresivo mayor (distimia, trastorno depresivo no especificado, duelo y trastorno adaptativo con estado de ánimo depresivo o mixto), los índices de validez diagnóstica fueron igualmente cuestionables (véase la Tabla 7.46).

Tampoco se modificó este patrón de eficacia diagnóstica cuestionable de las puntuaciones TB 75 y TB 85 de las escalas CC y D de la adaptación española del MCMI-III cuando el criterio de referencia de asignación a las categorías diagnósticas que se utilizó fue el diagnóstico resultante de la aplicación de una entrevista diagnóstica estructurada como la SCID-I, en lugar del diagnóstico basado en el juicio que los clínicos reflejaban en una ficha diagnóstica tras 3-4 sesiones de evaluación psicológica. De hecho, se volvió a encontrar para las puntuaciones TB 75 y 85 de las escalas del MCMI-III una sensibilidad baja (muy inferior al 70%), lo que indica de nuevo que muchos pacientes con trastorno depresivo mayor no puntúan por encima de las puntuaciones TB 75 y TB 85, es decir, se mantienen por tanto las dificultades de dichas puntuaciones para detectar correctamente a los pacientes con trastorno depresivo mayor.

La única excepción a este patrón general de rendimiento diagnóstico cuestionable fueron los resultados obtenidos con la puntuación TB 75 de la escala de Distimia cuando dicho rendimiento se evaluó utilizando como criterio de referencia el diagnóstico proporcionado por la SCID-I. En este caso, tanto el índice kappa como el índice de Youden sí superaron los valores estándar (índice kappa  $> 0.40$  e índice de Youden  $> 40$ ), aunque con una sensibilidad ligeramente inferior al 70% (63%). De hecho, tanto el poder o valor predictivo positivo como el poder o valor predictivo negativo de la puntuación TB 75 de la escala de Distimia superaron con creces el valor de 50% (70% y 79%, respectivamente), de manera que se podría estimar que la probabilidad de que un paciente que haya superado la TB 75 en la escala de Distimia tenga efectivamente un trastorno depresivo mayor sería del 70%, mientras que la probabilidad de que un paciente que puntúe por debajo de la TB 75 no padezca efectivamente un trastorno depresivo mayor sería del 79%, en ambos casos mayor que la probabilidad que supondría lanzar simplemente una moneda para decidir si un paciente padece o no dicho trastorno.

Curiosamente, esta excepción positiva en la eficacia del MCMI-III para el diagnóstico del trastorno depresivo mayor no se produjo en la escala teóricamente más relacionada con dicho trastorno, la escala de Depresión Mayor, sino en la escala de Distimia. Lamentablemente, no es posible comparar este último resultado con los resultados previos de la literatura científica ya que los estudios publicados hasta la fecha tan sólo han examinado la eficacia para el diagnóstico del trastorno depresivo mayor de la escala de Depresión Mayor del MCMI-III, pero no de su escala de Distimia (Millon, 1994; Millon et al., 1997; Hsu, 2000; Rossi y Sloore, 2005), por lo que sería deseable que futuras investigaciones prestaran mayor atención a la capacidad de las puntuaciones TB 75 y TB 85 de la escala de Distimia del MCMI-III para diagnosticar un trastorno depresivo mayor.

**Tabla 7.46.** Índices de validez para el diagnóstico del trastorno depresivo mayor obtenidos en este estudio para las escalas de Depresión Mayor y Distimia del MCMI-III

Índices	Diagnóstico de trastorno depresivo mayor frente a otros trastornos mentales (ficha clínica)				Diagnóstico de trastorno depresivo mayor frente a otros trastornos mentales no depresivos (ficha clínica)				Diagnóstico de trastorno depresivo mayor frente a otros trastornos mentales (SCID-I)			
	Depresión		Distimia		Depresión		Distimia		Depresión		Distimia	
	TB75	TB85	TB75	TB85	TB75	TB85	TB75	TB85	TB75	TB85	TB75	TB85
Prevalencia	20.7	20.7	20.7	20.7	25.5	25.5	25.5	25.5	37.5	37.5	37.5	37.5
Sensibilidad	31	12	54	24	31	12	54	24	29	7	63	21
Especificidad	93	100	74	95	96	100	79	94	97	100	83	98
VPP	54	54	35	55	72	100	47	59	87	100	70	88
VPN	83	100	86	82	80	77	83	78	69	64	79	67
Eficacia	80	81	80	70	79	77	72	76	71	65	76	69
Kappa	0.29	0.18	0.24	0.24	0.33	0.17	0.32	0.23	0.3	0.09	0.48	0.24
Youden	24	12	28	19	27	12	33	18	26	7	47	20

*Nota.* VPP: valor o poder predictivo positivo. VPN: valor o poder predictivo negativo. Eficacia: índice de eficacia global o porcentaje de pacientes correctamente clasificados (también denominado precisión, índice de validez o de eficiencia, o poder o valor diagnóstico total).



Por otro lado, en comparación con los resultados de los estudios previos que han analizado la eficacia para el diagnóstico del trastorno depresivo mayor de las puntuaciones TB 75 o TB 85 de la escala de Depresión Mayor del MCMI-III (Millon, 1994; Millon et al., 1997; Hsu, 2000; Rossi y Sloore, 2005), el patrón de resultados del presente estudio está en línea con el encontrado en los estudios de los manuales originales del MCMI-III en el sentido de que se encontró una mayor especificidad que sensibilidad y un mayor poder predictivo negativo que poder predictivo positivo (Millon, 1994; Millon et al., 1997; véase la Tabla 7.1). Sin embargo, los valores de sensibilidad y kappa del punto de corte TB 85 encontrados en el presente estudio son notablemente más bajos que los publicados en el primer manual definitivo del MCMI-III (Millon, 1994) y en la segunda edición del manual del MCMI-III (Millon et al., 1997; según los cálculos posteriores de Hsu, 2002).

No obstante, hay que recordar que en la investigación de la segunda edición del manual del MCMI-III los diagnósticos se realizaron sin utilizar necesariamente una entrevista diagnóstica estructurada como parte de la evaluación y, además, los clínicos debían rellenar el diagnóstico del paciente en la misma hoja en la que debían hacer constar las puntuaciones en el MCMI-III y podían incluir aquellos pacientes a los que ya hubieran aplicado el inventario, aunque conocieran sus puntuaciones, siempre que no las recordaran nítidamente (Millon et al., 1997). Este procedimiento podría haber hecho que se sobrestimara la validez diagnóstica de las escalas del MCMI-III.

El otro estudio publicado sobre las propiedades diagnósticas de las puntuaciones TB 75 o TB 85 de la escala de Depresión Mayor del MCMI-III respecto al diagnóstico del trastorno depresivo mayor, fue el realizado para estudiar su adaptación belga en flamencoparlantes (Rossi y Sloore, 2005; véase la tabla 7.1). En este caso, se utilizó como criterio diagnóstico de referencia un diagnóstico realizado a partir de la metodología

LEAD o evaluación longitudinal por expertos usando todos los datos disponibles (*Longitudinal, Expert, All Data*; Spitzer, 1983). Con el objetivo de asignar las puntuaciones TB 75 y TB 85 a las puntuaciones directas que mejores datos de eficacia diagnóstica ofrecían, en el estudio de Rossi y Sloore (2005) se llevó a cabo un proceso de selección de puntos de corte distinto al realizado por Millon y su equipo de investigación en la versión original del cuestionario. Rossi y Sloore (2005) seleccionaron puntos de corte con una sensibilidad por encima de 70%, a costa de tener una especificidad menor (51% y 55% para las puntuaciones TB 75 y TB 85, respectivamente). Es decir que, al contrario que en el presente estudio, en la adaptación belga los puntos de corte de TB 75 y TB 85 de la escala de Depresión Mayor del MCMI-III priman la detección del trastorno depresivo mayor, aún a costa de disminuir la capacidad del instrumento para detectar la ausencia del trastorno depresivo mayor en personas sin trastornos. No obstante, en dicha adaptación belga se obtuvieron unos índices kappa con valores similares a los encontrados en el presente estudio (véanse las Tabla 7.1 y Tabla 7.46), indicando una concordancia diagnóstica pobre o débil para las puntuaciones TB 75 y TB 85 de la escala de Depresión Mayor ( $< 40$ ; Altman, 1991), y, además, tras calcular los respectivos índices de Youden, éstos también fueron similares a los encontrados en el presente estudio, es decir, también fueron menores de 40 y no reflejaban un equilibrio adecuado entre sensibilidad y especificidad.

**¿Cuál es la validez diagnóstica de las puntuaciones TB 75 y TB 85 de las escalas CC y D del MCMI-III para el diagnóstico de cualquier trastorno depresivo?**

Coherentemente con los resultados obtenidos respecto a la eficacia para el diagnóstico del trastorno depresivo mayor, los resultados del presente estudio indican que, para el diagnóstico de cualquier trastorno depresivo, la validez diagnóstica de las puntuaciones TB 75 y TB 85 de las escalas CC y D de la adaptación española del MCMI-

III es, en general, también cuestionable. Para mayor claridad, se han resumido esos resultados en la Tabla 7.47 y de los mismos se puede concluir que, en general: (1) los índices de concordancia diagnóstica (kappa) de las puntuaciones TB 75 y TB 85 de las escalas CC y D se situaban entre débiles ( $< 0.20$ ) y pobres ( $< 0.40$ ); (2) los valores de sensibilidad de dichas puntuaciones fueron significativamente inferiores al estándar considerado como aceptable ( $< 70\%$ ), y (3) los índices de Youden no reflejaban un equilibrio adecuado entre sensibilidad y especificidad ( $< 40$ ).

Al igual que ocurría respecto al diagnóstico del trastorno depresivo mayor, los resultados del presente estudio también indican, respecto al diagnóstico de cualquier trastorno depresivo, que las puntuaciones TB 75 y TB 85 de las escalas de Depresión Mayor (CC) y Distimia (D) de la adaptación española del MCMI-III obtuvieron valores de especificidad elevados, superiores a 70% (e incluso superiores al 90% en el caso de las puntuaciones de la escala CC) e índices de poder o valor predictivo positivo y de poder o valor predictivo negativo que superaban el 60%. Por tanto, según estos datos, especialmente los datos de especificidad y valor predictivo negativo, los puntos de corte TB 75 y TB 85 de las escalas de CC y D de la adaptación española del MCMI-III serían útiles para detectar correctamente a los pacientes que no tienen un trastorno depresivo de cualquier tipo. Sin embargo, la excesivamente baja sensibilidad indica que muchos de los pacientes que sufren un trastorno depresivo de cualquier tipo puntúan por debajo de esos puntos de corte y, por tanto, no serían detectados utilizando las puntuaciones TB 75 y TB 85 de las escalas CC y D del MCMI-III.

Estos resultados están, en general, en línea con el patrón de resultados encontrado en los dos estudios previos realizados sobre la eficacia diagnóstica del MCMI-III para diagnosticar el conjunto de los trastornos depresivos (Gibeau y Choca, 2005; Saulsman, 2011; véase la Tabla 7.47).

**Tabla 7.47.** Índices de validez diagnóstica para el diagnóstico de cualquier trastorno depresivo de las escalas de Depresión Mayor y Distimia del MCMI-III en muestras de pacientes con trastornos psicológicos

Índices	Este estudio <sup>a</sup>				Gibeau y Choca (2005) <sup>b</sup>				Saulsman (2011) <sup>c</sup>			
	Depresión		Distimia		Depresión		Distimia		Depresión		Distimia	
	TB75	TB85	TB75	TB85	TB75	TB85	TB75	TB85	TB75	TB85	TB75	TB85
Prevalencia	39.4	39.4	39.4	39.4	42	42	42	42	58	58	58	58
Sensibilidad	23	6	48	13	45	26	73	32	55	35	66	21
Especificidad	96	100	78	94	79	90	58	86	87	91	64	95
VPP	79	100	60	61	60	67	56	62	85	84	72	85
VPN	65	62	70	62	67	63	75	63	58	50	58	46
Eficacia	67	63	66	62	65	63	65	63	—	—	—	—
Kappa	0.22	0.07	0.28	0.09	—	—	—	—	—	—	—	—
Youden <sup>d</sup>	19	6	27	8	24	16	31	18	42	26	30	16

*Nota.* VPP: valor o poder predictivo positivo. VPN: valor o poder predictivo negativo. Eficacia: índice de eficacia global o porcentaje de pacientes correctamente clasificados (también denominado precisión, índice de validez o de eficiencia, o poder o valor diagnóstico total).

<sup>a</sup>Eficacia para el diagnóstico de cualquier trastorno depresivo (trastorno depresivo mayor, distímico, adaptativo con estado de ánimo deprimido y depresivo no especificado, y duelo) sobre una muestra de pacientes ambulatorios españoles.

<sup>b</sup>Eficacia para el diagnóstico de cualquier trastorno depresivo (trastorno depresivo mayor, distímico, depresivo no especificado) sobre una muestra de pacientes veteranos de guerra estadounidenses (mayoritariamente hombres y tratados ambulatoriamente).

<sup>c</sup>Eficacia para el diagnóstico de cualquier trastorno depresivo (trastorno depresivo mayor y distimia) frente al grupo control de pacientes con diagnóstico de trastorno de ansiedad sobre una muestra de pacientes ambulatorios australianos sin comorbilidad.

<sup>d</sup>Este índice no aparece recogido en la publicación correspondiente, sino que ha sido calculado para este estudio a partir de los datos de la tabla.

Así, por ejemplo, en el estudio de Gibeau y Choca (2005) y utilizando como criterio de referencia el diagnóstico realizado por los clínicos sin utilizar necesariamente ninguna entrevista diagnóstica estructurada, las puntuaciones TB 75 y TB 85 de la escala de Depresión Mayor y la puntuación TB 85 de la escala de Distimia mostraron sensibilidades menores de 50% y especificidades mayores de 70%, de modo similar a lo encontrado en el presente estudio. También de manera similar al presente estudio, el valor predictivo positivo y el valor predictivo negativo de dichas puntuaciones para diagnosticar cualquier trastorno depresivo fueron superiores al 50%. La única discrepancia notable fue que la puntuación TB 75 de la escala de Distimia mostró en el estudio de Gibeau y Choca (2005), al contrario que en este estudio, una sensibilidad (73%) mayor del mínimo recomendado como aceptable y, sin embargo, una especificidad (58%) menor a dicho mínimo aceptable.

Por su parte, en el estudio de Saulsman (2011), en el cual se utilizó como criterio diagnóstico de referencia el proporcionado por la entrevista diagnóstica estructurada MINI, también se encontraron resultados similares, aunque ligeramente mejores, a los obtenidos en el presente estudio para las puntuaciones TB 75 y TB 85 de las escalas de Depresión Mayor y Distimia del MCMI-III (véase la tabla 7.47). En concreto, en esa investigación las puntuaciones TB 75 y TB 85 de dichas escalas mostraron sensibilidades inferiores a 70%, mientras que sus especificidades fueron superiores a 70% (salvo para la puntuación TB 75 de la escala de Distimia, cuya especificidad fue 64%). Además, el poder predictivo positivo se situó por encima del valor predictivo negativo, pero ambos, en cualquier caso, por encima del límite inferior de 50% recomendado para un instrumento diagnóstico (salvo para la puntuación TB 85 de la escala de Distimia, cuyo valor predictivo negativo fue 46%). Finalmente, cuando se calcularon los índices de Youden a partir de los datos de sensibilidad y especificidad ofrecidos por Saulsman (2011; véase la Tabla 4.47), dicho índices se situaron por debajo del umbral mínimo recomendado para

este estadístico (40), salvo en el caso de la puntuación TB 75 de la escala de Depresión Mayor, cuyo índice Youden era 42, pero que, no obstante, implicaba una sensibilidad (55%) por debajo de lo aceptable (70%).

Lamentablemente, ni el estudio de Gibeau y Choca (2005) ni en el de Saulsman (2011) se informó sobre el índice kappa, por lo que no es posible constatar si los débiles o pobres niveles de concordancia diagnóstica encontrados en el presente estudio son generalizables a otras investigaciones previas.

**¿Existen otras puntuaciones de las escalas CC y D del MCMI-III que tengan mejores índices de validez diagnóstica que sus puntuaciones de corte habituales (TB 75 y TB 85)?**

Como se ha comentado en los dos epígrafes anteriores, los resultados del presente estudio indican que las puntuaciones TB 75 y TB 85 de las escalas de Depresión Mayor (CC) y Distimia (D) de la adaptación española del MCMI-III priman, a la hora de diagnosticar el trastorno depresivo mayor o cualquier trastorno depresivo, la especificidad sobre la sensibilidad, de manera que si bien su validez para diagnosticar la ausencia de trastornos depresivos en pacientes que efectivamente tienen otros trastornos mentales sería aceptable, su validez para diagnosticar la presencia de trastornos depresivos en pacientes que efectivamente sufren esos trastornos no sería adecuada. Dado que esta última característica es tan importante en cualquier instrumento diagnóstico y dado que la concordancia diagnóstica de dichas puntuaciones con los criterios de referencia utilizados en este estudio se sitúa entre débil y pobre, la validez diagnóstica general de esas puntuaciones debe considerarse, en el mejor de los casos, como cuestionable.

Sin embargo, los resultados del presente estudio también indican que puede haber otras puntuaciones de las escalas de Depresión Mayor y Distimia de la adaptación española del MCMI-III que muestren mejores índices de eficacia diagnóstica, ya que, de hecho, esos resultados también indican que, en general, estas escalas son instrumentos con una eficacia diagnóstica adecuada para diagnosticar el trastorno depresivo mayor o cualquier trastorno depresivo.

Efectivamente, los resultados del presente estudio muestran, no sólo que las áreas bajo la curva ROC de las puntuaciones directas y de TB de las escalas de Depresión Mayor (CC) y Distimia (D) de la adaptación española del MCMI-III fueron significativamente mayores que el área mínima de 0.5 que representa un diagnóstico basado en el azar, sino que, además, dichas áreas superaron en todos los casos el criterio mínimo propuesto por algunos autores para valorar la exactitud diagnóstica de un instrumento ( $> 0.70$ ; Swets, 1988) y, en muchos casos (11 de 16), el criterio más estricto recomendado por otros ( $> 0.75$ ; Fan et al., 2006). En concreto, las áreas bajo la curva ROC de las puntuaciones directas y de TB de la escala CC para el diagnóstico del trastorno depresivo mayor según la ficha clínica, el área bajo la curva ROC de las puntuaciones TB de la escala CC para el diagnóstico de cualquier trastorno depresivo según la ficha clínica, las áreas bajo la curva ROC de las puntuaciones directas y de TB de las escalas CC y D para el diagnóstico del trastorno depresivo mayor con otros trastornos depresivos excluidos y según la ficha clínica, y las áreas bajo la curva ROC de las puntuaciones directas y de TB de las escalas CC y D para el diagnóstico del trastorno depresivo mayor según la SCID-I, mostraron todas ellas valores mayores de 0.75, los cuales son considerados clínicamente significativos (Fan et al., 2006). Es más, en el caso de estas últimas áreas bajo la curva ROC, las que se obtuvieron considerando como criterio diagnóstico de referencia la SCID-I, sus valores fueron todos ellos superiores a 0.80.

En consecuencia, cabría sospechar que otras puntuaciones de las escalas CC y D de la adaptación española del MCMI-III podrían tener índices de eficacia diagnóstica aceptables y, efectivamente, esto fue lo que también se encontró en el presente estudio. En concreto, tanto para diagnosticar el trastorno depresivo mayor como para diagnosticar cualquier trastorno depresivo, y tanto utilizando como criterio diagnóstico de referencia el juicio clínico basado en al menos 3-4 sesiones de evaluación como la información proporcionada por la entrevista diagnóstica estructurada SCID-I, en el presente estudio fue posible encontrar otros puntos de corte alternativos con sensibilidades, índices de kappa e índices de Youden ligeramente superiores a los que obtuvieron las puntuaciones de TB 75 y TB 85 y que en algunos casos superaban los valores mínimos recomendados para estos estadísticos. Es más, cuando se utilizó como criterio el diagnóstico de trastorno depresivo mayor de la entrevista SCID-I, otros puntos de corte alternativos a las puntuaciones TB 75 y TB 85 mostraron índices de eficacia diagnóstica notablemente mayores, con índices de concordancia diagnóstica (kappa) moderados (0.40-0.60; Altman, 1991), índices de Youden superiores al valor mínimo recomendado de 40 y sensibilidades y especificidades que simultáneamente eran superiores al valor mínimo de 70%.

Por ejemplo, para el diagnóstico del trastorno depresivo mayor y utilizando como criterio la ficha diagnóstica, la puntuación TB 75 de las escalas de CC y D del MCMI-III mostró el mejor índice kappa, pero hubo puntos de corte alternativos en la escala de CC como las puntuaciones TB 61 y TB 62 que tuvieron el mismo índice kappa, pero con un mayor y aceptable índice de Youden ( $> 40$ ). Las puntuaciones TB 61 y TB 62 primaron la sensibilidad ( $> 70\%$ ) por encima de la especificidad y tenían un poder predictivo positivo y un poder predictivo negativo superiores a 50%.



También se encontraron algunos puntos de corte alternativos con índices de eficacia diagnóstica para el diagnóstico del trastorno depresivo mayor ligeramente superiores a las puntuaciones TB 75 y TB 85 cuando el análisis de la eficacia diagnóstica se realizó excluyendo los pacientes con un diagnóstico depresivo distinto del trastorno depresivo mayor. Por ejemplo, los puntos de corte alternativos TB 61 y TB 62 de la escala CC tuvieron un índice de Youden superior a 40, con un índice kappa moderado (0.40) y con un mejor equilibrio entre sensibilidad y especificidad, aunque el poder predictivo positivo fue ligeramente inferior al 50%. Otras puntuaciones como la PD 11 de las escalas de CC y D del MCMI-III también obtuvieron índices diagnósticos ligeramente mejores que las puntuaciones TB 75, con índices kappa moderados (0.40) e índices de Youden superiores a 40. En el caso de la escala D, el punto de corte TB 66 mostró un mayor índice de Youden que las puntuaciones TB 75 y 85 (ligeramente superiores a 40), aunque con un índice kappa débil, similar a la puntuación TB 75. Así mismo, la PD 11 tuvo un índice de Youden más elevado que los puntos de corte originales, así como un índice kappa ligeramente superior, aunque también débil.

Cuando se evaluó la capacidad para diagnosticar cualquier trastorno depresivo, también se encontraron puntos de corte alternativos a TB 75 y TB 85 que tenían mejores índices kappa, aunque de nuevo débiles, y mayores índices Youden (aunque inferiores al criterio mínimo recomendado de 40). Por ejemplo, las puntuaciones TB 61 y TB 62 tuvieron, en el caso de la escala CC, una sensibilidad y especificidad muy cercanas a 70%, con un índice de Youden cercano a 40 y un kappa de 0.37. La puntuación directa 8 de la escala de CC y la puntuación directa 9 en la escala de D tuvieron índices de eficacia diagnóstica similares a las puntuaciones TB 61 y TB 62 y, por tanto, ligeramente superiores a las puntuaciones TB 75 y TB 85.

La cosecha de puntos de corte alternativos a TB 75 y TB 85 se acentuó de manera significativa cuando se utilizó la SCID-I como criterio para el diagnóstico del trastorno depresivo mayor. En este caso, las puntuaciones alternativas TB 61 y TB 62 para las escalas de CC y D tuvieron índices de validez diagnóstica, en general, aceptables ( $\kappa > 0.40$ , índice de Youden  $> 40$ , sensibilidad  $> 70\%$ , poder predictivo positivo  $> 50\%$ , poder predictivo negativo  $> 50\%$ ). En el caso de la escala de CC, estos puntos de corte alternativos también tuvieron especificidades superiores a  $70\%$ , aunque en el caso de la escala de D se situó ligeramente por debajo. Además, en la línea de lo comentado para las puntuaciones TB 61 y TB 62, algunas puntuaciones directas también tuvieron índices de eficacia diagnóstica adecuados como la puntuación PD 9 de la escala de CC y la PD 10 de la escala de D ( $\kappa > 0.40$ , índice de Youden  $> 40$ , sensibilidad  $> 70\%$ , especificidad  $> 70\%$ , poder predictivo positivo  $> 50\%$ , poder predictivo negativo  $> 50\%$ ).

En resumen, los resultados del presente estudio sugieren que: (a) existen puntuaciones directas y de TB de las escalas de Depresión Mayor y Distimia de la adaptación española del MCMI-III que pueden mejorar la capacidad de dichas escalas para diagnosticar un trastorno depresivo mayor o cualquier trastorno depresivo y que se configuran como mejores puntos de corte que las puntuaciones TB 75 y TB 85, y (b) que, en general, las escalas de Depresión Mayor y Distimia de la adaptación española del MCMI-III son instrumentos aceptables para el diagnóstico del trastorno depresivo mayor o cualquier trastorno depresivo siempre y cuando se utilice un punto de corte con índices diagnósticos adecuados.

Estos resultados, pues, confirman las afirmaciones de Hsu (2002) y de Rossi y Sloore (2005) y son consistentes con los resultados de estos últimos investigadores en el sentido de que la metodología utilizada para calcular las puntuaciones TB no garantiza de

manera automática que la selección de los puntos de corte sea la más eficaz desde el punto de vista diagnóstico, sino que son necesarios estudios específicos de eficacia diagnóstica para comprobar qué punto de corte es el más adecuado para cada población.

Por otro lado, los resultados del presente estudio también mostraron que, cuando se utiliza la SCID-I como criterio de referencia del diagnóstico de trastorno depresivo mayor, se encontraron estadísticos de eficacia diagnóstica más adecuados para ambas escalas depresivas del MCMI-III, superiores por lo general a los encontrados cuando se utilizaba como criterio de referencia el diagnóstico reflejado en la ficha diagnóstica. Sobre este punto se volverá más adelante, pero baste señalar por ahora que estos datos corroboran también las ideas de Ganellen (1996) sobre que la metodología elegida por los investigadores en el cálculo de los índices de eficacia diagnóstica tiene una gran influencia a la hora de establecer el rendimiento diagnóstico de los instrumentos.

**¿Son mejores los índices de validez diagnóstica de las escalas CC y D del MCMI-III que los de otros instrumentos ampliamente utilizados para evaluar la depresión como, por ejemplo, el BDI-II?**

Los resultados del presente estudio indican que los índices de validez diagnóstica de las escalas CC y D del MCMI-III para el diagnóstico del trastorno depresivo mayor o para el diagnóstico de cualquier trastorno depresivo no son mejores que los índices del BDI-II, sino más bien similares.

En el presente estudio, tanto las áreas bajo la curva ROC de las escalas CC y D del MCMI-III como las áreas bajo la curva del BDI-II superaron el valor mínimo que representa una exactitud aceptable ( $> 0.70$ ; Swets, 1988) y en algunos casos superaron las exactitudes que otros investigadores más rigurosos consideran necesarias para que un instrumento tenga significación clínica ( $> 0.75$ ; Fan et al., 2006), pero no se

encontraron, ni para el diagnóstico del trastorno depresivo mayor ni para el conjunto de los trastornos depresivos, diferencias estadísticamente significativas entre las escalas del MCMI-III y el BDI-II o entre las propias escalas del MCMI-III, ya que los intervalos de confianza de sus áreas bajo la curva ROC se solapaba entre sí y, además, todos los estadísticos de comparación dos a dos de áreas bajo la curva ROC confirmaron que no existían diferencias estadísticamente significativas entre dichas áreas.

De hecho, los índices de eficacia diagnóstica (sensibilidad, especificidad, kappa, índice de Youden, valor predictivo positivo, valor predictivo negativo) para diversas puntuaciones del BDI-II fueron muy similares a los obtenidos para los mejores puntos de corte de las escalas de Depresión Mayor y Distimia del MCMI-III. Así, por ejemplo, los índices kappa máximos obtenidos por los puntos de corte en las tres escalas fueron muy similares en cada uno de los análisis realizado. De hecho, al igual que en el caso de las escalas de Depresión Mayor y Distimia del MCMI-III, cuando se utilizó como criterio diagnóstico de referencia el diagnóstico de trastorno depresivo mayor reflejado en la entrevista SCID-I, se encontró para el BDI puntos de corte con adecuados estadísticos de eficacia diagnóstica (kappa > 0.40, índice de Youden > 40, sensibilidad >70%, poder predictivo positivo > 50%, poder predictivo negativo > 50%) como, por ejemplo, las puntuaciones de 22, 23 y 24.

Los resultados sobre la eficacia diagnóstica del BDI-II que se obtuvieron en el presente estudio fueron ligeramente mejores que los obtenidos en el primer estudio de esta tesis doctoral. En concreto, por ejemplo, las áreas bajo la curva ROC del BDI en el primer estudio fueron 0.77 y 0.782, ambas estadísticamente significativas, para el diagnóstico de trastorno depresivo mayor, y 0.689, no estadísticamente significativa, para el diagnóstico de cualquier trastorno depresivo, mientras que en el presente estudio las áreas para el diagnóstico de trastorno depresivo mayor fueron de 0.748, 0.778 y 0.86,

todas ellas estadísticamente significativas, y 0.737, también estadísticamente significativa, para el diagnóstico de cualquier trastorno depresivo. De hecho, en el estudio primero todas las puntuaciones del BDI-II evaluadas mostraron índices de concordancia diagnóstica (kappa) por debajo del estándar de 0.40 e índices de Youden por debajo del estándar de 40, y ninguna fue capaz de presentar simultáneamente valores adecuados de sensibilidad y especificidad, es decir, por encima de 70% ambos. Sin embargo, en este segundo estudio, se pudieron identificar varias puntuaciones de corte del BDI-II con índices de concordancia diagnóstica aceptables ( $\text{kappa} > 0.40$ ), con índices de Youden que reflejaban un equilibrio adecuado entre sensibilidad y especificidad ( $> 40$ ) y con sensibilidades y especificidades simultáneas por encima del estándar de 70%.

Puesto que las muestras de ambos estudios proceden de la misma población (los pacientes con trastornos psicológicos atendidos en la Clínica Universitaria de Psicología de la UCM), cabría pensar que esa ligera mejora en los índices de validez diagnóstica del BDI-II podría deberse, además de a las variaciones naturales de las muestras, al mayor rigor metodológico empleado en este segundo estudio respecto al primero en el sentido de que se amplió el número de sesiones de evaluación que los clínicos tenían que realizar antes de llegar a un diagnóstico y de que este diagnóstico se realizaba después de que el clínico hubiera aplicado, al menos, una entrevista diagnóstica estructurada (la SCID-I). Es más, en esa dirección apunta el hecho de que los mejores índices de validez diagnóstica se obtuvieron en el presente estudio cuando se utilizó como criterio de referencia el diagnóstico proporcionado por la SCID-I en lugar del diagnóstico que los clínicos habían reflejado en la ficha diagnóstica de los pacientes, dirección a la que también apuntan los resultados del estudio de Sanz y García-Vera (2013).

Como se comentó en el capítulo anterior, la muestra de pacientes que participó en el estudio de Sanz y García-Vera (2013) se solapaba en parte con la que participó en el

presente estudio. Sin embargo, para el análisis del rendimiento diagnóstico del BDI-II, en el estudio de Sanz y García-Vera (2013) se seleccionaron aquellos pacientes a los que se les había aplicado, además de la SCID-I, un instrumento de heteroevaluación para medir la presencia de síntomas depresivos y valorar la gravedad de la depresión (el Listado para el Episodio Depresivo o LED; Sanz et al., 2005) y sólo se analizó la capacidad del BDI-II para detectar un trastorno depresivo mayor frente a un trastorno no depresivo. Por tanto, en ese estudio se descartaron todos aquellos pacientes que: a) no tenían el LED, o b) presentaban un diagnóstico principal de trastorno distímico, trastorno depresivo no especificado, trastorno adaptativo con estado de ánimo depresivo, trastorno adaptativo mixto con ansiedad y estado de ánimo depresivo o trastorno adaptativo con alteración mixta de las emociones y el comportamiento, o c) presentaban como diagnóstico secundario un trastorno depresivo o un trastorno adaptativo de los subtipos mencionados anteriormente. Por otro lado, en el estudio de Sanz y García-Vera (2013) el diagnóstico de referencia fue el anotado por el psicólogo en la ficha diagnóstica sobre la base, al menos, de la información de la SCID-I, pero también de la lista de cotejo de síntomas depresivos (LED). Por tanto, no es de extrañar que en comparación con los resultados obtenidos en el presente estudio cuando se utilizó como criterio de referencia el diagnóstico que los clínicos habían reflejado en la ficha diagnóstica de los pacientes, Sanz y García-Vera (2013) informaran de índices de validez diagnóstica ligeramente superiores. Así, por ejemplo, el área bajo la curva ROC del BDI-II fue 0.83 en el estudio de Sanz y García-Vera (2013), mientras que en el presente estudio, para el diagnóstico de trastorno depresivo mayor y tras excluir los pacientes con otros trastornos depresivos, el área bajo la curva ROC fue de 0.778. Sin embargo, cuando en el presente estudio se utilizó como criterio de referencia el diagnóstico proporcionado por la SCID-I, el área de la curva ROC que mostró el BDI-II para el diagnóstico del trastorno depresivo mayor fue ligeramente superior al informado por Sanz y García-Vera (2013) (0.86 frente a 0.83).

En definitiva, estos resultados subrayan la importancia que tiene la elección del criterio diagnóstico de referencia en los estudios de validación diagnóstica de los instrumentos. De hecho, en el presente estudio se han encontrado algunas discrepancias entre el diagnóstico consignado en la ficha diagnóstica y el diagnóstico registrado en la aplicación del módulo de trastornos del estado de ánimo de la SCID-I y, aunque no significativamente, los índices de validez diagnóstica tanto de las escalas CC y D del MCMI-III como del BDI-II fueron mejores cuando se utilizó como criterio de referencia el diagnóstico proporcionado por la SCID-I que cuando se utilizó el diagnóstico consignado en la ficha diagnóstica.

Existen varias explicaciones, no excluyentes, que podrían dar cuenta de esas discrepancias. En primer lugar, habría que recordar las dificultades que presenta la realización de un diagnóstico adecuado en psicopatología y, en especial, respecto a los trastornos depresivos, dificultades que ya fueron comentadas en parte en el capítulo segundo de esta tesis doctoral. En este sentido, habría que destacar en especial la gran similitud, comorbilidad y continuidad entre la sintomatología del trastorno depresivo mayor y la de los distintos trastornos y problemas que forman el grupo más amplio de trastornos depresivos o problemas relacionados con la depresión (p. ej., distimia, trastorno depresivo no especificado, duelo, trastorno adaptativo con estado de ánimo deprimido), así como la gran similitud y comorbilidad que también guarda la sintomatología del trastorno depresivo mayor con la de los trastornos de ansiedad, todo lo cual dificulta la realización del diagnóstico diferencial. En segundo lugar, habitualmente el intervalo entre la aplicación de la SCID-I y la anotación del diagnóstico en la ficha era de dos sesiones, con un intervalo de una semana entre cada una de ellas, durante las cuales los psicólogos podrían no recordar el diagnóstico registrado en la SCID-I o haber considerado que otra etiqueta diagnóstica era más adecuada para recoger la sintomatología del paciente.

En cualquier caso, los datos sugieren que cuando en este estudio segundo se utilizó como criterio de referencia el diagnóstico de la ficha clínica, las áreas bajo la curva ROC para el diagnóstico del trastorno depresivo mayor fueron muy similares a las del estudio primero y los estadísticos de eficacia diagnóstica de los distintos puntos de corte tan solo fueron ligeramente mejores a los del estudio primero y se situaban, en general, por debajo de los valores considerados estándar. Sin embargo, cuando en este estudio segundo se utilizó como criterio el diagnóstico de trastorno depresivo mayor proporcionado por la SCID-I, en estos momentos uno de los instrumentos y procedimientos considerados el “patrón oro” del diagnóstico en psicopatología, efectivamente las áreas de la curva ROC y los estadísticos de eficacia diagnóstica de algunos puntos de corte del BDI-II y de las escalas CC y D del MCMI-III fueron superiores.

Dado que, como se señalaba en el capítulo anterior, existe un alto porcentaje de pacientes no diagnosticados con trastornos depresivos que sí son detectados y diagnosticados con dichos trastornos cuando los clínicos realizan los diagnósticos mediante entrevistas diagnósticas estructuradas (Rogers, 2003) y dada la amplia literatura que, en general, apoya la mayor validez diagnóstica de la SCID-I respecto a las entrevistas clínicas habituales (Basco et al., 2000; Fenning et al., 1994; Fenning et al., 1996; Kranzler, et al., 1995; Kranzler et al., 1996), los resultados del presente estudio también subrayan la necesidad de utilizar entrevistas diagnósticas estructuradas a la hora de establecer el criterio diagnóstico de referencia en los estudios de validación diagnóstica del BDI-II, del MCMI-III y, en general, de otros instrumentos psicopatológicos.

### **Limitaciones**

Por supuesto, las conclusiones, indicaciones y sugerencias que se han presentado en los epígrafes anteriores tienen que valorarse en el contexto de las



limitaciones del presente estudio. Es cierto que en el mismo se han adoptado mejoras metodológicas respecto a la mayoría de los estudios previos sobre la validez diagnóstica de las escalas CC y D del MCMI-III. En este sentido conviene recordar que en los propios estudios realizados para la segunda edición del manual original del MCMI-III (Millon et al., 1997) así como en otras investigaciones (Gibeau y Choca, 2005), los diagnósticos de referencia se realizaron sin utilizar necesariamente una entrevista diagnóstica estructurada como parte del proceso de evaluación. En otros estudios (Saulsman, 2011), en cambio, se utilizó una entrevista diagnóstica estructurada abreviada (*Mini-International Neuropsychiatric Interview* o MINI) cuyo estándar de referencia es precisamente la entrevista diagnóstica estructurada utilizada en el presente estudio, la SCID-I (Sheehan et al., 1997). Igualmente, en el diseño de este estudio se ha pretendido controlar en lo posible, mediante la realización de diagnósticos ciegos a las puntuaciones del MCMI-III y del BDI-II, las variables contaminadoras presentes en otros estudios sobre la validez del MCMI-III, ya que, por ejemplo, en los propios estudios realizados para la segunda edición del manual original del MCMI-III (Millon et al., 1997) los clínicos debían rellenar el diagnóstico del paciente en la misma hoja en la que debían hacer constar las puntuaciones en el MCMI-III, pudiendo, además, incluir pacientes a los que ya hubieran aplicado el inventario, aunque conocieran sus puntuaciones, siempre y cuando no las recordaran nítidamente. Finalmente, los resultados del estudio también indican que la fiabilidad o validez convergente del MCMI-III en la muestra concreta de pacientes del presente estudio o las propias características de ésta no parecen que hubiesen sido una fuente importante de sesgos y errores en la medida que los índices de fiabilidad y de validez convergente obtenidos fueron aceptables y similares a los de la literatura científica y, así mismo, la frecuencia de síntomas y trastornos depresivos en la muestra de pacientes también fue similar a la informada en la literatura científica.

Pero, precisamente, una limitación importante de esta investigación tiene que ver con la fiabilidad del diagnóstico de referencia realizado por los psicólogos, ya que no se puso en marcha ningún procedimiento que permitiera estimar dicha fiabilidad, de modo que el criterio sobre el que se puso a prueba la validez diagnóstica del MCMI-III podría ponerse en duda. Futuras investigaciones, pues, podrían mejorar la metodología del presente estudio llevando a cabo un procedimiento para medir la fiabilidad interjueces del diagnóstico. De hecho, otra opción de mejora en el procedimiento de este estudio sería poner en marcha un entrenamiento más exhaustivo sobre la aplicación de la SCID-I, lo cual a la postre redundaría en una mejora en la fiabilidad del diagnóstico de referencia y de su validez.

La validez también podría verse incrementada en futuras investigaciones si, además de utilizar una entrevista diagnóstica estructurada, se utiliza algún otro procedimiento de los que hoy en día se consideran el “patrón de oro” del diagnóstico como, por ejemplo, la metodología LEAD de evaluación longitudinal por varios expertos usando todos los datos disponibles (Spitzer, 1983) o cualquier otro procedimiento similar (véase Aboraya, France, Young, Curci y LePage, (2005). De hecho, en un estudio previo sobre la validez diagnóstica de las escalas de depresión del MCMI-III (Rossi y Sloore, 2005) se utilizó como criterio de referencia el diagnóstico clínico alcanzado tras una evaluación clínica realizada aplicando la metodología LEAD, aunque desafortunadamente no se incluyó como parte de dicha metodología la aplicación de una entrevista diagnóstica estructurada.

## 7.5. Conclusiones

Este es el primer estudio publicado sobre la validez diagnóstica de las escalas de Depresión Mayor (CC) y Distimia (D) de la adaptación española del MCMI-III, por lo que son necesarios nuevos estudios de replicación para comprobar la generalizabilidad de los resultados obtenidos sobre la validez diagnóstica de dichas escalas. No obstante, los datos encontrados sobre el rendimiento diagnóstico tanto de esas escalas del MCMI-III como del BDI-II señalan que los tres son instrumentos útiles para el diagnóstico del trastorno depresivo mayor o del conjunto de los trastornos depresivos siempre y cuando se utilicen puntos de corte adecuados y se tengan en cuenta las limitaciones de dichos puntos de corte. Así, aunque los puntos de corte TB 75 y TB 85 de las escalas CC y D del MCMI-III no alcanzaron los estándares mínimos aceptables para un instrumento diagnóstico, sí existían puntos de corte alternativos a los propuestos en el manual de la adaptación española del MCMI-III que en la presente muestra tenían mejor rendimiento diagnóstico, alcanzando de hecho los estándares mínimos recomendados para un instrumento diagnóstico cuando se utiliza como criterio diagnóstico el módulo de los trastornos del estado de ánimo de la SCID-I.

Especialmente se debería tener en cuenta que las puntuaciones TB 75 y TB 85, si bien podrían ayudar a detectar la ausencia de trastornos depresivos en pacientes con otros trastornos mentales, dejan sin embargo sin detectar un alto porcentaje de pacientes que sufren efectivamente de trastornos depresivos. En tanto en cuanto los puntos de corte más aconsejables para ser utilizados en el proceso diagnóstico son diferentes a los recogidos en el manual de la adaptación española del instrumento (TB 75 y TB 85), se recomienda utilizar las escalas CC o D del MCMI-III como instrumento diagnóstico de los trastornos depresivos junto con otras fuentes de información, a ser posible a través de la

aplicación de una entrevista diagnóstica estructurada, máxime si el principal objetivo de la evaluación psicológica es llegar a un diagnóstico psicopatológico.

# **Capítulo 8.**

# **CONCLUSIONES**

# **GENERALES**



## 8. Conclusiones generales

Como se comentó al inicio de esta tesis doctoral, los estudios epidemiológicos de las últimas décadas han confirmado que los trastornos depresivos tienen una alta prevalencia, son muy incapacitantes y producen un alto coste social (Ebmeier, Donaghey y Steele, 2006), por lo que su detección tiene en la actualidad un papel relevante en psicopatología. Durante los últimos 30 años la actividad diagnóstica de los trastornos mentales ha crecido espectacularmente y los sistemas diagnósticos como el del DSM-IV están siendo ampliamente utilizados. Pero hasta ahora no se han encontrado marcadores biológicos o criterios causales inequívocos que permitan llegar a un diagnóstico totalmente fiable y válido de los trastornos depresivos, de manera que el “patrón oro” para su detección es la aplicación de una entrevista diagnóstica estructurada por un clínico con experiencia. Como ya se comentaba en la revisión teórica del capítulo 2 de esta tesis, “los endeble fundamentos de los sistemas diagnósticos no los hacen irrelevantes”, sino que los mismos “son cada vez más utilizados, y su uso tiene enormes repercusiones económicas y sociales” (Vázquez, Sánchez y Romero, 2011).

Se ha señalado que los profesionales que utilizan entrevistas no estructuradas para la realización de este proceso diagnóstico, las cuales son las más habituales en la práctica clínica cotidiana, dejan sin detectar un alto porcentaje de pacientes con trastornos depresivos (Roger, 2003). En este sentido, recientemente se han publicado los resultados de los estudios de fiabilidad interjueces de las categorías diagnósticas que formarán el futuro DSM-5. En el caso del trastorno depresivo mayor y utilizando diagnósticos independientes y ciegos, el índice kappa se situaba dentro del rango de fiabilidad “cuestionable” (kappa entre 0.20 y 0.40) (Regier et. al., 2013).

Además de las entrevistas clínicas (estructuradas o no estructuradas), los profesionales e investigadores disponen de cuestionarios, inventarios y escalas válidos

para la valoración de la sintomatología depresiva, algunos de ellos diseñados específicamente para el diagnóstico del trastorno depresivo mayor, con una alta correlación entre ellos y que han ido confluyendo en torno a los criterios de los manuales DSM. Entre tales cuestionarios, inventarios y escalas destaca el Inventario Clínico Multiaxial de Millon o MCMI.

El MCMI, en sus distintas versiones (MCMI, MCMI-II y MCMI-III), ha sido recomendado para su uso en labores diagnósticas tanto en la práctica clínica como en la investigación y en este sentido está siendo utilizado de manera amplia internacionalmente y en España. Los manuales del MCMI-II y MCMI-III señalan que el diagnóstico es uno de sus objetivos principales (Millon, 2002; Millon, Davis y Millon, 2007). Pero, para determinar la utilidad de un cuestionario como instrumento diagnóstico se deben conocer sus propiedades de validez diagnóstica en la población con la que va a ser utilizado y para los puntos de corte que se van emplear en la interpretación de sus puntuaciones, y no se deben asumir automáticamente los datos de validez obtenidos con otras poblaciones diferentes y otros puntos de corte, incluyendo los obtenidos con las versiones originales de un instrumento (ITC, 2000).

No se han publicado hasta ahora estudios de validez diagnóstica de las escalas de las adaptaciones españolas del MCMI. Además, respecto a las versiones originales del MCMI, existen dudas metodológicas sobre los estudios que han evaluado sus propiedades psicométricas así como un déficit en los estadísticos publicados, de forma que la validez diagnóstica de este inventario ha sido objeto de una gran controversia (Hsu, 2002; Retzlaff, 1996; Grove y Vrieze, 2009). Por lo tanto, el objetivo general de esta tesis doctoral fue analizar las propiedades diagnósticas de las escalas de Depresión Mayor (CC) y Distimia (D) del MCMI-II y del MCMI-III para la detección de los trastornos depresivos en muestras españolas de pacientes con trastornos mentales y, de forma



secundaria, comparar dichas propiedades con las del Inventario de Depresión de Beck-II (BDI-II), quizás el instrumento más utilizado y validado para la evaluación de la presencia y gravedad de sintomatología depresiva.

En este contexto se encuadran los dos estudios empíricos independientes que se incluyen en esta tesis doctoral y que versan sobre la validez diagnóstica, para el diagnóstico del trastorno depresivo mayor y del conjunto de los trastornos depresivos, de las escalas de CC y D de las adaptaciones españolas del MCMI-II y del MCMI-III. El primer estudio se centró en el MCMI-II, tuvo un diseño retrospectivo instrumental y se realizó con una muestra ambulatoria de 181 pacientes con diversos trastornos mentales. Los evaluadores conocían generalmente las puntuaciones de los pacientes tanto del MCMI-II como del BDI-II antes de realizar el diagnóstico del paciente, de manera que no fue posible descartar la existencia de sesgos o de contaminación cruzada entre las puntuaciones de estos inventarios y los diagnósticos realizados por los psicólogos, y, además, dichos evaluadores realizaron los diagnósticos sin utilizar necesariamente una entrevista diagnóstica estructurada.

Estos dos aspectos metodológicos fueron corregidos en el segundo estudio, el cual respondía a un diseño instrumental prospectivo y en cual se analizó la validez diagnóstica del MCMI-III con una muestra de 203 pacientes ambulatorios con diversos trastornos mentales. En este estudio, se aplicaron los instrumentos a validar (MCMI-III y BDI-II) antes de que los evaluadores llegaran a un diagnóstico, y éste se basó en un procedimiento de evaluación psicológica que implicaba al menos 3-4 entrevistas clínicas y la aplicación del módulo de los trastornos del estado de ánimo de la entrevista diagnóstica estructurada SCID-I (First et al., 1999). Este diagnóstico se registraba en una ficha diagnóstica y los análisis de eficacia diagnóstica del MCMI-III y del BDI-II se

realizaron teniendo en cuenta, como criterio de referencia, por un lado el diagnóstico de esa ficha diagnóstica o clínica y, por otro lado, el resultado proporcionado por la SCID-I.

Las principales conclusiones que se pueden extraer de los resultados de ambos estudios serían las siguientes:

1. La validez diagnóstica de las puntuaciones TB 75 y TB 85 de las escalas de Depresión Mayor y Distimia de las adaptaciones españolas del MCMI-II y del MCMI-III, y que representan los puntos de corte habituales del MCMI, es cuestionable, ya que la mayoría de sus índices de rendimiento diagnóstico se encuentran por debajo de los estándares aceptables para un instrumento diagnóstico, tanto para el diagnóstico del trastorno depresivo mayor como para el diagnóstico del conjunto de los trastornos depresivos. Específicamente, para las puntuaciones TB 75 y TB 85 de ambas escalas, los índices de sensibilidad se encuentran por debajo del estándar de 70%, los índices de concordancia diagnóstica (kappa) están por debajo de 0.40, y los índices de Youden no reflejan un equilibrio adecuado entre sensibilidad y especificidad ( $< 40$ ). Tan sólo los índices de especificidad ( $> 70\%$ ) y de valor predictivo negativo ( $> 50\%$ ) apoyan de forma consistente la validez diagnóstica de las escalas CC y D del MCMI-II y del MCMI-III para el diagnóstico del trastorno depresivo mayor. En resumen, se debería tener en cuenta que las puntuaciones TB 75 y TB 85 de las escalas CC y D de las adaptaciones españolas del MCMI-II y del MCMI-III dejan sin detectar un alto porcentaje de pacientes con trastornos depresivos y no cumplen la mayoría de los estándares para una eficacia diagnóstica aceptable, aunque podrían ser útiles para descartar la presencia de un trastorno depresivo.
2. Existen otras puntuaciones directas y de TB en el MCMI-II y en el MCMI-III que muestran índices de rendimiento diagnóstico mejores que las puntuaciones TB 75

y TB 85 y que podrían considerarse como puntos de corte alternativos. En consecuencia, la estrategia metodológica de utilizar las puntuaciones de TB para obtener los puntos de corte de un instrumento diagnóstico no siempre asegura que se obtengan los mejores puntos de corte y que dichos puntos muestren índices de validez diagnóstica adecuados. Sería conveniente, por tanto, utilizar otras estrategias metodológicas que examinen el rendimiento diagnóstico de un amplio conjunto de puntuaciones (p. ej., el análisis de la curva ROC) para poder seleccionar entre ellas la que muestren los mejores índices de validez diagnóstica y, por supuesto, confirmar la generalizabilidad de dichos índices en estudios de replicación.

3. En el caso del MCMI-III, los puntos de corte alternativos no sólo obtienen índices de validez diagnóstica ligeramente superiores a los de las puntuaciones TB 75 y TB 85, sino que llegan a alcanzar los estándares mínimos recomendados para un instrumento diagnóstico ( $\kappa > 0.40$ , índice de Youden  $> 40$ , sensibilidad  $> 70\%$ , especificidad  $> 70\%$ , poder o valor predictivo positivo  $> 50\%$ , poder o valor predictivo negativo  $> 50\%$ ), especialmente cuando se utiliza como criterio diagnóstico de referencia el obtenido a través una entrevista diagnóstica estructurada, la SCID-I.
4. Este último hecho, junto con el dato de que las áreas bajo la curva ROC de las puntuaciones directas y de TB de las escalas CC y D del MCMI-III superan en todos los casos el criterio mínimo propuesto por algunos autores para valorar la exactitud diagnóstica de un instrumento ( $> 0.70$ ) y, en muchos casos, el criterio más estricto recomendado por otros ( $> 0.75$ ), sugiere que los dos escalas de la adaptación española del MCMI-III son instrumentos útiles para el diagnóstico del trastorno depresivo mayor o del conjunto de los trastornos depresivos siempre y

cuando se utilicen puntos de corte adecuados y se tengan en cuenta las limitaciones de dichos puntos de corte.

5. En el caso del MCMI-II, aunque otros puntos de corte alternativos obtienen índices de validez diagnóstica similares o ligeramente superiores a las puntuaciones TB 75 y TB 85, dichos índices no superan en general los valores aceptables para el diagnóstico del conjunto de los trastornos depresivos, lo cual, unido al dato de que las áreas bajo la curva ROC de las puntuaciones directas y de TB de las escalas CC y D del MCMI-II para el diagnóstico del conjunto de los trastornos depresivos no superan en ningún caso el criterio mínimo de exactitud diagnóstica de un instrumento ( $> 0.70$ ), sugiere que dichas escalas no son aceptables para realizar ese diagnóstico.
6. En cuanto a la capacidad del MCMI-II para diagnosticar el trastorno depresivo mayor, aunque otros puntos de corte alternativos obtienen índices de validez diagnóstica similares o ligeramente superiores a las puntuaciones TB 75 y TB 85, dichos índices tampoco superan en general los valores aceptables para un instrumento diagnóstico. No obstante, puesto que las áreas bajo la curva ROC de las puntuaciones directas y de TB de las escalas CC y D del MCMI-II para el diagnóstico del trastorno depresivo mayor sí superan en la mayoría de los casos el criterio mínimo de exactitud diagnóstica de un instrumento ( $> 0.70$ ), cabría plantearse que la adecuación de dichas escalas para realizar el diagnóstico del trastorno depresivo mayor necesitaría nuevas investigaciones, máxime cuando en el estudio de validez del MCMI-II no se pudo contar con una entrevista diagnóstica estructurada que proporcionara el criterio diagnóstico de referencia y, de hecho, los datos de esta tesis sugieren que la utilización de dicha entrevista mejoró notablemente los índices de eficacia diagnóstica de las escalas del MCMI-III,

probablemente porque mejoró precisamente la propia validez del criterio diagnóstico de referencia.

7. Aunque teóricamente la escala de Depresión Mayor (CC) del MCMI-II o del MCMI-III debería mostrar un mejor rendimiento para el diagnóstico del trastorno depresivo mayor que la escala de Distimia (D) del MCMI-II o del MCMI-III, no existen diferencias significativas entre las escalas CC y las escalas D en cuanto a su índices de eficacia diagnóstica para realizar dicho diagnóstico. Es más, tampoco se encontraron diferencias significativas en el rendimiento diagnóstico entre las escalas CC y las escalas D para el diagnóstico del conjunto de los trastornos depresivos. Por tanto, no parecen abordar adecuadamente el diagnóstico diferencial entre trastorno depresivo mayor y distimia, lo cual, por otro lado, no es de extrañar dado que las escalas CC y las escalas D tienen un gran solapamiento de ítems y no valoran adecuadamente los criterios temporales del DSM para ambos trastornos. Como Saulsman (2011) sugiere, ambas escalas pueden estar proporcionando información redundante y, por tanto, “requieren una investigación más a fondo” (p. 81).
8. Los índices de validez diagnóstica de las escalas CC y D del MCMI-II y del MCMI-III para el diagnóstico del trastorno depresivo mayor o para el diagnóstico de cualquier trastorno depresivo no son mejores, sino similares, a los índices de validez diagnóstica que obtiene el BDI-II. De hecho, cuando se utiliza como criterio diagnóstico de referencia el obtenido a través una entrevista diagnóstica estructurada, existen varios puntos de corte en el BDI-II que obtienen índices de validez diagnóstica que alcanzan los estándares mínimos recomendados para un instrumento diagnóstico ( $\kappa > 0.40$ , índice de Youden  $> 40$ , sensibilidad  $>$

70%, especificidad > 70%, poder o valor predictivo positivo > 50%, poder o valor predictivo negativo > 50%).

Por supuesto, las conclusiones anteriores se deben valorar teniendo en cuenta las limitaciones de los estudios de la presente tesis doctoral. Por ejemplo, en el caso del estudio sobre el MCMI-III no se realizó ningún estudio sobre la fiabilidad de los diagnósticos, mientras que en el caso del estudio del MCMI-III, además de esa misma limitación, se podría señalar que el diagnóstico no se realizó a partir de una entrevista estructurada y que los evaluadores conocían generalmente las puntuaciones de los cuestionarios antes de realizar el diagnóstico.

Por otro lado, las dudas planteadas sobre la validez diagnóstica de las puntuaciones TB 75 y TB 85 de las escalas de Depresión Mayor y Distimia del MCMI-II y del MCMI-III no afectan a la utilización de dichas escalas como instrumentos para detectar síntomas y síndromes depresivos y medir su gravedad. El mismo proceso de construcción de los instrumentos, el propio contenido de sus ítems y los datos de fiabilidad, validez factorial y validez convergente que presentan los manuales de las adaptaciones españolas avalan ese uso, lo que obviamente es muy importante para evaluar, por ejemplo, el estado general de los pacientes o la efectividad de un proceso terapéutico. De hecho, en la presente tesis doctoral también se han encontrado datos que confirman la utilidad de las escalas de Depresión Mayor y Distimia del MCMI-III para la valoración de la sintomatología depresiva, ya que se ha encontrado una alta correlación de dichas escalas con el BDI-II, instrumento de referencia para la valoración de la gravedad de la sintomatología depresiva.

En resumen, uno de los principales objetivos del MCMI-II y del MCMI-III es “diagnosticar los porcentajes de los pacientes que realmente presentan trastornos en los ámbitos clínicos” (Millon, Davis y Millon, 2007, pp. 20-21) y en este sentido está siendo

utilizado. Sin embargo, los usuarios de las adaptaciones españolas del MCMI-II y del MCMI-III deberían tener en cuenta que el sistema de puntuaciones TB que utilizan ambos instrumentos para establecer los puntos de corte que indican la presencia de un trastorno, no garantiza automáticamente la eficacia diagnóstica de sus escalas. A partir de los resultados de los dos estudios empíricos de esta tesis doctoral y de los resultados de los estudios realizados con anterioridad sobre la eficacia diagnóstica de las escalas de los trastornos depresivos del MCMI, lo más recomendable sería no utilizar las escalas de Depresión Mayor y Distimia de las adaptaciones españolas del MCMI-II y del MCMI-III como instrumentos principales en el diagnóstico de los trastornos depresivos, al menos en tanto en cuanto no se realicen estudios posteriores que confirmen los resultados positivos encontrados en esta investigación y refuten también sus resultados negativos. En todo caso, de los instrumentos analizados en esta investigación y sobre la base de los resultados obtenidos, parece que las escalas de Depresión Mayor y Distimia del MCMI-III y el BDI-II serían los instrumentos más útiles para llegar a un diagnóstico de trastorno depresivo mayor o de un trastorno depresivo en general, pero siempre y cuando se utilicen puntos de corte adecuados y se tengan en cuenta las limitaciones de dichos puntos de corte. En este sentido, consideramos adecuadas las propias palabras del manual del MCMI-III cuando señala que "como instrumento de autoinforme, el MCMI sólo representa una faceta de la evaluación total del paciente. La riqueza y precisión de cualquier medida de autoinforme mejora cuando sus resultados se evalúan en el contexto de otras fuentes clínicas" (Millon, Davis y Millon, 2007, p. 23), y en este sentido creemos que se debería utilizar.

# **ANEXOS**





**Anexo 1.** Criterios diagnósticos para el episodio depresivo mayor según criterios DSM-IV-TR (APA, 2002)

- A) Estar presentes 5 o más de los criterios sobre un listado de 9, durante un periodo de al menos dos semanas consecutivas, que representa un cambio con respecto a la actividad previa. Además, al menos uno de los síntomas debe ser el criterio (1) o (2):
- (1) estado de ánimo depresivo.
  - (2) disminución acusada del interés o de la capacidad para el placer.
  - (3) pérdida o aumento de peso importante de peso o de apetito.
  - (4) insomnio o hipersomnia.
  - (5) agitación o enlentecimiento psicomotores.
  - (6) fatiga o pérdida de energía.
  - (7) sentimiento de culpa o inutilidad excesiva.
  - (8) disminución de la capacidad para pensar o concentrarse.
  - (9) pensamiento recurrente de muerte o ideación suicida.
- B) Los síntomas no deben cumplir criterios para un episodio mixto.
- C) Los síntomas provocan malestar clínicamente significativo o deterioro social, laboral o de otras áreas importantes de la actividad del individuo.
- D) Los síntomas no son debidos a los efectos fisiológicos directos de una sustancia o una enfermedad médica.
- E) Los síntomas no se explican mejor por la presencia de un duelo (p. ej., después de la pérdida de un ser querido), los síntomas persisten durante más de 2 meses o se caracterizan por una acusada incapacidad funcional, preocupaciones mórbidas de inutilidad, ideación suicida, síntomas psicóticos o enlentecimiento psicomotor.

**Anexo 2.** Criterios diagnósticos para el episodio depresivo mayor según los la CIE-10 (OMS, 1993)

En los episodios depresivos típicos el enfermo que las padece sufre un humor depresivo, una pérdida de la capacidad de interesarse y disfrutar de las cosas, una disminución de su vitalidad que lleva a una reducción de su nivel de actividad y a un cansancio exagerado, que aparece incluso tras un esfuerzo mínimo. También son manifestaciones de los episodios depresivos:

- a) La disminución de la atención y concentración.
- b) La pérdida de la confianza en sí mismo y sentimientos de inferioridad.
- c) Las ideas de culpa y de ser inútil (incluso en las episodios leves).
- d) Una perspectiva sombría del futuro.
- e) Los pensamientos y actos suicidas o de autoagresiones.
- f) Los trastornos del sueño.
- g) La pérdida del apetito.

En el caso del **episodio depresivo leve** se den cumplir dos de los tres criterios principales y al menos dos criterios de entre los restantes. Para el **episodio depresivo moderado** además de dos de los tres criterios principales, cuatro los criterios de la lista. En el caso del **episodio depresivo grave**, deberán cumplirse los tres criterios iniciales, además de al menos cuatro criterios de entre los restantes.

**Anexo 3.** Criterios diagnósticos para el trastorno depresivo mayor de episodio único según criterios DSM-IV-TR (APA, 2002)

- A) Presencia de un único episodio depresivo mayor para el trastorno depresivo de episodio único, más de uno si es recidivante
- B) El episodio depresivo mayor no se explica mejor por la presencia de un trastorno esquizoafectivo y no está superpuesto a una esquizofrenia, un trastorno esquizofreniforme, un trastorno delirante o un trastorno psicótico no especificado.
- C) Nunca se ha producido un episodio maníaco, un episodio mixto o un episodio hipomaníaco. Nota: Esta exclusión no es aplicable si todos los episodios similares a la manía, a los episodios mixtos o a la hipomanía son inducidos por sustancias o por tratamientos o si se deben a los efectos fisiológicos directos de una enfermedad médica.

**Anexo 4.** Criterios diagnósticos para el trastorno distímico según criterios DSM-IV-TR (APA, 2002)

- A) Estado de ánimo crónicamente depresivo la mayor parte del día de la mayoría de los días, manifestado por el sujeto u observado por los demás, durante al menos 2 años.
- B) Presencia, mientras está deprimido, de dos (o más) de los siguientes síntomas:
  - (1) pérdida o aumento de apetito
  - (2) insomnio o hipersomnia
  - (3) falta de energía o fatiga
  - (4) baja autoestima
  - (5) dificultades para concentrarse o para tomar decisiones
  - (6) sentimientos de desesperanza
- C) Durante el período de 2 años (1 año en niños y adolescentes) de la alteración, el sujeto no ha estado sin síntomas de los Criterios A y B durante más de 2 meses seguidos.
- D) No ha habido ningún episodio depresivo mayor durante los primeros 2 años de la alteración
- E) Nunca ha habido un episodio maníaco, un episodio mixto o un episodio hipomaníaco y nunca se han cumplido los criterios para el trastorno ciclotímico.
- F) La alteración no aparece exclusivamente en el transcurso de un trastorno psicótico crónico, como son la esquizofrenia o el trastorno delirante.
- G) Los síntomas no son debidos a los efectos fisiológicos directos de una sustancia (p. ej., una droga, un medicamento) o a enfermedad médica (p. ej., hipotiroidismo).
- H) Los síntomas causan un malestar clínicamente significativo o deterioro social, laboral o de otras áreas importantes de la actividad del individuo.

**Anexo 5.** Escalas clínicas de personalidad, escalas de personalidad de patología grave, escalas de síndromes clínicos y escalas de síndromes graves del MCMI-II y el MCMI-III (Millon, 2002, adaptación española de A. Ávila-Espada.; Millon, Davis y Millon, 2007, adaptación española de V. Cardenal y M. P. Sánchez)

Patrones clínicos de personalidad		Síndromes clínicos	
<b>1</b>	Esquizoide	<b>A</b>	Trastorno de ansiedad
<b>2 A</b>	Evitativo	<b>H</b>	Trastorno somatomorfo
<b>2 B</b>	Depresivo <sup>1</sup>	<b>N</b>	Trastorno bipolar
<b>3</b>	Dependiente	<b>D</b>	Trastorno distímico
<b>4</b>	Histriónico	<b>B</b>	Dependencia de alcohol
<b>5</b>	Narcisista	<b>T</b>	Dependencia de sustancias
<b>6 A</b>	Antisocial	<b>R</b>	Trastorno de estrés postraumático <sup>1</sup>
<b>6 B</b>	Agresivo (sádico)		
<b>7</b>	Compulsivo		
<b>8 A</b>	Negativista (pasivo-agresivo)		
<b>8 B</b>	Autodestructivo		
Patología grave de personalidad		Síndromes clínicos graves	
<b>S</b>	Esquizotípico	<b>SS</b>	Trastorno del pensamiento
<b>C</b>	Límite	<b>CC</b>	Depresión mayor
<b>P</b>	Paranoide	<b>PP</b>	Trastorno delirante
		Índices modificadores	
		<b>X</b>	Sinceridad
		<b>Y</b>	Deseabilidad Social
		<b>Z</b>	Devaluación
		<b>V</b>	Validez

HISTORIA Nº:

TERAPEUTA:

FECHA:

## Anexo 6. FICHA DIAGNÓSTICA

**Sexo:** ☐ Varón ☐ Mujer **Edad:** \_\_\_\_\_

**Estado civil:** ☐ Soltero/a ☐ Casado/a o conviviendo de forma estable con una pareja  
☐ Viudo/a ☐ Separado/a o divorciado/a

**Estudios:** ☐ No ha ido a la escuela ☐ Estudios primarios  
☐ EGB o equivalente ☐ FP, Bachillerato, BUP o equivalente  
☐ Universitarios ☐ Otros estudios no reglados

**Situación laboral:** ☐ Trabaja ☐ Jubilado o pensionista ☐ Parado/a  
☐ Estudiante ☐ Sus labores ☐ Otra situación

**Profesión:** \_\_\_\_\_

### DEMANDA ASISTENCIAL

**Nº DE SESIÓN:** ☐ 3º entrevista ☐ 4º entrevista ☐ Otra: \_\_\_\_\_

### DIAGNÓSTICOS PROVISIONALES

**Grado de certeza: (0-10)**

Eje I: \_\_\_\_\_  
 \_\_\_\_\_  
 \_\_\_\_\_

☐ Sin diagnóstico ☐ Diagnóstico o estado aplazado

Eje II (Trastornos de personalidad):

\_\_\_\_\_  
 \_\_\_\_\_

☐ Sin diagnóstico ☐ Diagnóstico o estado aplazado

SI CREE QUE EL PACIENTE TIENE ALGÚN TIPO DE PROBLEMA DEPRESIVO (AUNQUE NO SE CORRESPONDA CON NINGÚN DIAGNÓSTICO), ¿CUÁL ESTIMA QUE ES LA GRAVEDAD DE SU DEPRESIÓN?

☐ Sin depresión ☐ Leve ☐ Moderada ☐ Grave

¿Qué medicación está tomando actualmente el paciente o ha tomado en las dos últimas semanas?

\_\_\_\_\_  
☐ Ninguna

**Anexo 7.** Comparación de curvas ROC de las puntuaciones directas de las escalas de Distimia (D) y Depresión Mayor (CC) del MCMI-II y del BDI-II con respecto al diagnóstico de trastorno depresivo mayor

Variable 1	Puntuación total del BDI-II
Variable 2	Puntuación directa de la escala Depresión Mayor
Variable 3	Puntuación directa de la escala Distimia
Variable de clasificación	Diagnóstico de trastorno depresivo mayor
Criterio seleccionado	1

	AUC	SE <sup>a</sup>	95% CI <sup>b</sup>
Puntuación total del BDI-II	0.771	0.0517	0.691 a 0.838
Puntuación directa de la escala Depresión Mayor	0.699	0.0664	0.614 a 0.774
Puntuación directa de la escala Distimia	0.742	0.0610	0.659 a 0.813

<sup>a</sup> DeLong et al., 1988

<sup>b</sup> Binomial exact

**Comparación de las curvas ROC por pares**

<b>BDI-II ~ Puntuación directa de la escala Depresión Mayor</b>	
Diferencia entre áreas	0.0720
Error Estándar <sup>c</sup>	0.0547
Intervalo de confianza al 95%	-0.0352 a 0.179
Estadístico z	1.316
Nivel de significación	P = 0.1882
<b>BDI-II ~ Puntuación directa de la escala Distimia</b>	
Diferencia entre áreas	0.0292
Error Estándar <sup>c</sup>	0.0568
Intervalo de confianza al 95%	-0.0821 a 0.140
Estadístico z	0.514
Nivel de significación	P = 0.6071
<b>Puntuación directa de la escala Depresión Mayor ~ Puntuación directa de la escala Distimia</b>	
Diferencia entre áreas	0.0428
Error Estándar <sup>c</sup>	0.0249
Intervalo de confianza al 95%	-0.00599 a 0.0917
Estadístico z	1.720
Nivel de significación	P = 0.0855

<sup>c</sup> DeLong et al., 1988



**Anexo 8.** Comparación de curvas ROC de las puntuaciones tasa base de las escalas de Distimia (D) y Depresión Mayor (CC) del MCMI-II y del BDI-II con respecto al diagnóstico de trastorno depresivo mayor

Variable 1	Puntuación total del BDI-II
Variable 2	Puntuación tasa base de la escala Depresión Mayor
Variable 3	Puntuación tasa base de la escala de Distimia
Variable de clasificación	Diagnóstico de trastorno depresivo mayor
Criterio seleccionado	1

	AUC	SE <sup>a</sup>	95% CI <sup>b</sup>
Puntuación total del BDI-II	0.740	0.0502	0.667 a 0.804
Puntuación tasa base de la escala Depresión Mayor	0.683	0.0667	0.607 a 0.752
Puntuación tasa base de la escala Distimia	0.724	0.0584	0.650 a 0.789

<sup>a</sup> DeLong et al., 1988

<sup>b</sup> Binomial exact

**Comparación de las curvas ROC por pares**

<b>BDI-II ~ Puntuación tasa base de la escala de Depresión Mayor</b>	
Diferencia entre áreas	0.0570
Error Estándar <sup>c</sup>	0.0529
Intervalo de confianza al 95%	-0.0467 a 0.161
Estadístico z	1.078
Nivel de significación	P = 0.2812
<b>BDI-II ~ Puntuación tasa base de la escala de Distimia</b>	
Diferencia entre áreas	0.0160
Error Estándar <sup>c</sup>	0.0488
Intervalo de confianza al 95%	-0.0797 a 0.112
Estadístico z	0.328
Nivel de significación	P = 0.7427
<b>Puntuación directa de la tasa base de Depresión Mayor ~ Puntuación directa de la tasa base de Distimia</b>	
Diferencia entre áreas	0.0410
Error Estándar <sup>c</sup>	0.0309
Intervalo de confianza al 95%	-0.0196 a 0.101
Estadístico z	1.326
Nivel de significación	P = 0.1849

<sup>c</sup> DeLong et al., 1988

**Anexo 9.** Comparación de curvas ROC de las puntuaciones directas de las escalas de Distimia (D) y de Depresión Mayor (CC) del MCMI-II y de las puntuaciones directas del BDI-II con respecto al diagnóstico del conjunto de trastornos depresivos

Variable 1	Puntuación total del BDI-II
Variable 2	Puntuación directa de la escala Depresión Mayor
Variable 3	Puntuación directa de la escala Distimia
Variable de clasificación	Diagnóstico del conjunto de trastornos depresivos
Criterio seleccionado	1

	AUC	SE <sup>a</sup>	95% CI <sup>b</sup>
Puntuación total del BDI-II	0.689	0.0458	0.604 a 0.765
Puntuación directa de la escala de Depresión Mayor	0.665	0.0477	0.579 a 0.744
Puntuación directa de la escala Distimia	0.679	0.0463	0.594 a 0.757

<sup>a</sup> DeLong et al., 1988

<sup>b</sup> Binomial exact

**Comparación de las curvas ROC por pares**

<b>BDI-II ~ Puntuación directa de la escala Depresión Mayor</b>	
Diferencia entre áreas	0.0235
Error Estándar <sup>c</sup>	0.0362
Intervalo de confianza al 95%	-0.0475 a 0.0944
Estadístico z	0.648
Nivel de significación	P = 0.5169
<b>BDI-II ~ Puntuación directa de la escala Distimia</b>	
Diferencia entre áreas	0.00951
Error Estándar <sup>c</sup>	0.0347
Intervalo de confianza al 95%	-0.0585 a 0.0775
Estadístico z	0.274
Nivel de significación	P = 0.7839
<b>Puntuación directa de la escala Depresión Mayor ~ Puntuación directa de la escala Distimia</b>	
Diferencia entre áreas	0.0140
Error Estándar <sup>c</sup>	0.0171
Intervalo de confianza al 95%	-0.0195 a 0.0474
Estadístico z	0.817
Nivel de significación	P = 0.4137

<sup>c</sup> DeLong et al., 1988

**Anexo 10.** Comparación de las curvas ROC de las puntuaciones tasa base (TB) de las escalas de Distimia (D) y de Depresión Mayor (CC) del MCMI-II y de las puntuaciones directas del BDI-II con respecto al diagnóstico del conjunto de trastornos depresivos

Variable 1	Puntuación total del BDI-II
Variable 2	Puntuación tasa base de la escala Depresión Mayor
Variable 3	Puntuación tasa base de la escala Distimia
Variable de clasificación	Diagnóstico del conjunto de trastornos depresivos
Criterio seleccionado	1

	AUC	SE <sup>a</sup>	95% CI <sup>b</sup>
Puntuación total del BDI-II	0.670	0.0420	0.594 a 0.740
Puntuación tasa base de la escala Depresión Mayor	0.640	0.0450	0.563 a 0.712
Puntuación tasa base de la escala Distimia	0.638	0.0440	0.562 a 0.710

<sup>a</sup> DeLong et al., 1988

<sup>b</sup> Binomial exact

#### Comparación de las curvas ROC por pares

<b>BDI-II ~ Puntuación tasa base de la escala Depresión Mayor</b>	
Diferencia entre áreas	0.0297
Error Estándar <sup>c</sup>	0.0359
Intervalo de confianza al 95%	-0.0406 a 0.100
Estadístico z	0.829
Nivel de significación	P = .4072
<b>BDI-II ~ Puntuación tasa base de la escala Distimia</b>	
Diferencia entre áreas	0.0312
Error Estándar <sup>c</sup>	0.0368
Intervalo de confianza al 95%	-0.0409 a 0.103
Estadístico z	0.848
Nivel de significación	P = 0.3966
<b>Puntuación tasa base de la escala Depresión Mayor ~ Puntuación tasa base de la escala Distimia</b>	
Diferencia entre áreas	0.00145
Error Estándar <sup>c</sup>	0.0232
Intervalo de confianza al 95%	-0.0440 a 0.0469
Estadístico z	0.0627
Nivel de significación	P = 0.9500

<sup>c</sup> DeLong et al., 1988

**Anexo 11.** Curvas ROC de las puntuaciones directas de la escala de Distimia y Depresión Mayor del MCMI-II con respecto al diagnóstico de Trastorno Depresivo Mayor (pacientes con el diagnóstico de otros trastornos depresivos excluidos)

Variable 1	Puntuación total del BDI-II
Variable 2	Puntuación directa de la escala Depresión Mayor
Variable 3	Puntuación directa de la escala Distimia
Classification variable Select	Diagnóstico de trastorno depresivo mayor 1

	AUC	SE <sup>a</sup>	95% CI <sup>b</sup>
Puntuación total del BDI-II	0.782	0.0499	0.693 a 0.855
Puntuación directa de la escala Depresión Mayor	0.718	0.0646	0.625 a 0.800
Puntuación directa de la escala Distimia	0.754	0.0583	0.662 a 0.831

<sup>a</sup> DeLong et al., 1988

<sup>b</sup> Binomial exact

#### Comparación de las curvas ROC por pares

<b>BDI-II ~ Puntuación directa de la escala Depresión Mayor</b>	
Diferencia entre áreas	0.0637
Error Estándar <sup>c</sup>	0.0514
Intervalo de confianza al 95%	-0.0370 a 0.164
Estadístico z	1.240
Nivel de significación	P = 0.2149
<b>BDI-II ~ Puntuación directa de la escala Distimia</b>	
Diferencia entre áreas	0.0284
Error Estándar <sup>c</sup>	0.0514
Intervalo de confianza al 95%	-0.0724 a 0.129
Estadístico z	0.552
Nivel de significación	P = 0.5809
<b>Puntuación directa de la escala Depresión Mayor ~ Puntuación directa de la escala Distimia</b>	
Diferencia entre áreas	0.0353
Error Estándar <sup>c</sup>	0.0242
Intervalo de confianza al 95%	-0.0121 a 0.0827
Estadístico z	1.461
Nivel de significación	P = 0.1439

<sup>c</sup> DeLong et al., 1988

**Anexo 12.** Comparación de curvas ROC de las puntuaciones tasa base de las escalas de Distimia y Depresión Mayor del MCMI-II con respecto al diagnóstico de trastorno depresivo mayor (pacientes con el diagnóstico de otros trastornos depresivos excluidos)

Variable 1	Puntuación total del BDI-II
Variable 2	Puntuación tasa base de la escala Depresión Mayor
Variable 3	Puntuación tasa base de la escala Distimia
Variable de clasificación	Diagnóstico de trastorno depresivo mayor
Criterio seleccionado	1

	AUC	SE <sup>a</sup>	95% CI <sup>b</sup>
Puntuación total del BDI-II	0.753	0.0489	0.673 a 0.822
Puntuación tasa base de la escala Depresión Mayor	0.696	0.0655	0.613 a 0.771
Puntuación tasa base de la escala Distimia	0.731	0.0571	0.650 a 0.803

<sup>a</sup> DeLong et al., 1988

<sup>b</sup> Binomial exact

#### Comparación de las curvas ROC por pares

<b>BDI-II ~ Puntuación tasa base de la escala Depresión Mayor</b>	
Diferencia entre áreas	0.0572
Error Estándar <sup>c</sup>	0.0506
Intervalo de confianza al 95%	-0.0420 a 0.156
Estadístico z	1.130
Nivel de significación	P = 0.2585
<b>BDI-II ~ Puntuación tasa base de la escala Distimia</b>	
Diferencia entre áreas	0.0220
Error Estándar <sup>c</sup>	0.0463
Intervalo de confianza al 95%	-0.0687 a 0.113
Estadístico z	0.475
Nivel de significación	P = 0.6345
<b>Puntuación tasa base de la escala Depresión Mayor ~ Puntuación tasa base de la escala Distimia</b>	
Diferencia entre áreas	0.0352
Error Estándar <sup>c</sup>	0.0297
Intervalo de confianza al 95%	-0.0229 a 0.0933
Estadístico z	1.187
Nivel de significación	P = 0.2351

<sup>c</sup> DeLong et al., 1988

**Anexo 13.** Comparación de curvas ROC de las puntuaciones directas y tasa base de las escalas de Distimia y Depresión Mayor del MCMI-III con respecto al diagnóstico de trastorno depresivo mayor

Variable 1	BDI-II
Variable 2	Puntuación directa de la escala Depresión Mayor
Variable 3	Puntuación tasa base de la escala Depresión Mayor
Variable 4	Puntuación directa de la escala Distimia
Variable 5	Puntuación tasa base de la escala Distimia
Variable de clasificación	Diagnóstico de trastorno depresivo mayor

	AUC	SE <sup>a</sup>	95% CI <sup>b</sup>
BDI- II	0.748	0.0417	0.681 a 0.807
Puntuación directa de la escala de Depresión Mayor	0.774	0.0391	0.709 a 0.831
Puntuación tasa base de la escala Depresión Mayor	0.748	0.0423	0.681 a 0.808
Puntuación directa de la escala Distimia	0.751	0.0419	0.684 a 0.810
Puntuación tasa base de la escala Distimia	0.723	0.0442	0.654 a 0.784

<sup>a</sup> DeLong et al., 1988

<sup>b</sup> Binomial exact

#### Comparación de las curvas ROC por pares

<b>BDI-II ~ Puntuación directa de la escala Depresión Mayor</b>	
Diferencia entre áreas	0.0262
Error Estándar <sup>c</sup>	0.0427
Intervalo de confianza al 95%	-0.0575 a 0.110
Estadístico z	0.614
Nivel de significación	P = 0.5392
<b>BDI-II ~ Puntuación tasa base de la escala Depresión Mayor</b>	
Diferencia entre áreas	0.000325
Error Estándar <sup>c</sup>	0.0441
Intervalo de confianza al 95%	-0.0862 a 0.0868
Estadístico z	0.00736
Nivel de significación	P = 0.9941
<b>BDI-II ~ Puntuación directa de la escala Distimia</b>	
Diferencia entre áreas	0.00300
Error Estándar <sup>c</sup>	0.0326
Intervalo de confianza al 95%	-0.0610 a 0.0670
Estadístico z	0.0920
Nivel de significación	P = 0.9267
<b>BDI-II ~ Puntuación tasa base de la escala Distimia</b>	
Diferencia entre áreas	0.0252
Error Estándar <sup>c</sup>	0.0312
Intervalo de confianza al 95%	-0.0359 a 0.0864
Estadístico z	0.809

Nivel de significación	P = 0.4186
Puntuación directa de la escala Depresión Mayor ~ Puntuación tasa base de la escala Depresión Mayor	
Diferencia entre áreas	0.0259
Error Estándar <sup>c</sup>	0.0245
Intervalo de confianza al 95%	-0.0221 a 0.0739
Estadístico z	1.058
Nivel de significación	P = 0.2902
Puntuación directa de la escala Depresión Mayor ~ Puntuación directa de la escala Distimia	
Diferencia entre áreas	0.0232
Error Estándar <sup>c</sup>	0.0309
Intervalo de confianza al 95%	-0.0374 a 0.0838
Estadístico z	0.751
Nivel de significación	P = 0.4525
Puntuación directa de la escala Depresión Mayor ~ Puntuación tasa base de la escala Distimia	
Diferencia entre áreas	0.0515
Error Estándar <sup>c</sup>	0.0387
Intervalo de confianza al 95%	-0.0244 a 0.127
Estadístico z	1.330
Nivel de significación	P = 0.1835
Puntuación tasa base de la escala Depresión Mayor ~ Puntuación directa de la escala Distimia	
Diferencia entre áreas	0.00268
Error Estándar <sup>c</sup>	0.0410
Intervalo de confianza al 95%	-0.0777 a 0.0831
Estadístico z	0.0653
Nivel de significación	P = 0.9480
Puntuación tasa base de la escala Depresión Mayor ~ Puntuación tasa base de la escala Distimia	
Diferencia entre áreas	0.0256
Error Estándar <sup>c</sup>	0.0356
Intervalo de confianza al 95%	-0.0442 a 0.0953
Estadístico z	0.718
Nivel de significación	P = 0.4725
Puntuación directa de la escala Distimia ~ Puntuación tasa base de la escala Distimia	
Diferencia entre áreas	0.0282
Error Estándar <sup>c</sup>	0.0242
Intervalo de confianza al 95%	-0.0191 a 0.0756
Estadístico z	1.170
Nivel de significación	P = 0.2422

<sup>c</sup> DeLong et al., 1988

**Anexo 14.** Comparación de curvas ROC de las puntuaciones directas y tasa base de las escalas de Distimia y Depresión Mayor del MCMI-III con respecto al diagnóstico del conjunto de los trastornos depresivos

Variable 1	BDI-II
Variable 2	Puntuación directa de la escala Depresión Mayor
Variable 3	Puntuación tasa base de la escala Depresión Mayor
Variable 4	Puntuación directa de la escala Distimia
Variable 5	Puntuación tasa base de la escala Distimia
Variable de clasificación	Diagnóstico de trastorno depresivo mayor

	AUC	SE <sup>a</sup>	95% CI <sup>b</sup>
BDI- II	0.737	0.0357	0.669 a 0.798
Puntuación directa de la escala de Depresión Mayor	0.748	0.0351	0.681 a 0.808
Puntuación tasa base de la escala Depresión Mayor	0.723	0.0365	0.654 a 0.785
Puntuación directa de la escala Distimia	0.743	0.0355	0.675 a 0.803
Puntuación tasa base de la escala Distimia	0.706	0.0374	0.637 a 0.769

<sup>a</sup> DeLong et al., 1988

<sup>b</sup> Binomial exact

#### Comparación de las curvas ROC por pares

<b>BDI-II ~ Puntuación directa de la escala de Depresión Mayor</b>	
Diferencia entre áreas	0.0110
Error Estándar <sup>c</sup>	0.0338
Intervalo de confianza al 95%	-0.0551 a 0.0772
Estadístico z	0.327
Nivel de significación	P = 0.7437
<b>BDI-II ~ Puntuación directa de la escala Distimia</b>	
Diferencia entre áreas	0.0145
Error Estándar <sup>c</sup>	0.0325
Intervalo de confianza al 95%	-0.0492 a 0.0781
Estadístico z	0.445
Nivel de significación	P = 0.6562
<b>BDI-II ~ Puntuación tasa base de la escala Depresión Mayor</b>	
Diferencia entre áreas	0.00571
Error Estándar <sup>c</sup>	0.0365
Intervalo de confianza al 95%	-0.0657 a 0.0772
Estadístico z	0.157
Nivel de significación	P = 0.8755
<b>BDI-II ~ Puntuación tasa base de la escala Distimia</b>	
Diferencia entre áreas	0.0311
Error Estándar <sup>c</sup>	0.0331
Intervalo de confianza al 95%	-0.0338 a 0.0960
Estadístico z	0.939
Nivel de significación	P = 0.3479



Puntuación directa de la escala Depresión Mayor ~ Puntuación directa de la escala Distimia	
Diferencia entre áreas	0.0255
Error Estándar <sup>c</sup>	0.0262
Intervalo de confianza al 95%	-0.0259 a 0.0769
Estadístico z	0.972
Nivel de significación	P = 0.3310
Puntuación directa de la escala Depresión Mayor ~ Puntuación tasa base de la escala Depresión Mayor	
Diferencia entre áreas	0.00532
Error Estándar <sup>c</sup>	0.0160
Intervalo de confianza al 95%	-0.0260 a 0.0367
Estadístico z	0.333
Nivel de significación	P = 0.7392
Puntuación directa de la escala Depresión Mayor ~ Puntuación tasa base de la escala Distimia	
Diferencia entre áreas	0.0421
Error Estándar <sup>c</sup>	0.0303
Intervalo de confianza al 95%	-0.0173 a 0.102
Estadístico z	1.388
Nivel de significación	P = 0.1651
Puntuación directa de la escala Distimia ~ Puntuación tasa base de la escala Depresión Mayor	
Diferencia entre áreas	0.0202
Error Estándar <sup>c</sup>	0.0295
Intervalo de confianza al 95%	-0.0377 a 0.0780
Estadístico z	0.683
Nivel de significación	P = 0.4943
Puntuación directa de la escala Distimia ~ Puntuación tasa base de la escala Distimia	
Diferencia entre áreas	0.0166
Error Estándar <sup>c</sup>	0.0143
Intervalo de confianza al 95%	-0.0114 a 0.0447
Estadístico z	1.162
Nivel de significación	P = 0.2451
Puntuación tasa base de la escala Depresión Mayor ~ Puntuación tasa base de la escala Distimia	
Diferencia entre áreas	0.0368
Error Estándar <sup>c</sup>	0.0286
Intervalo de confianza al 95%	-0.0192 a 0.0928
Estadístico z	1.288
Nivel de significación	P = 0.1978

<sup>c</sup> DeLong et al., 1988

**Anexo 15.** Comparación de curvas ROC de las puntuaciones directas y tasa base de las escalas de Distimia y Depresión Mayor del MCMI-III con respecto al diagnóstico de trastorno depresivo mayor (pacientes con el diagnóstico de otros trastornos depresivos excluidos)

Variable 1	BDI-II
Variable 2	Puntuación directa de la escala Depresión Mayor
Variable 3	Puntuación tasa base de la escala Depresión Mayor
Variable 4	Puntuación directa de la escala Distimia
Variable 5	Puntuación tasa base de la escala Distimia
Variable de clasificación	Diagnóstico de trastorno depresivo mayor

	AUC	SE <sup>a</sup>	95% CI <sup>b</sup>
BDI- II	0.774	0.0407	0.701 a 0.837
Puntuación directa de la escala de Depresión Mayor	0.796	0.0385	0.724 a 0.855
Puntuación tasa base de la escala Depresión Mayor	0.797	0.0388	0.726 a 0.857
Puntuación directa de la escala Distimia	0.768	0.0412	0.695 a 0.831
Puntuación tasa base de la escala Distimia	0.762	0.0416	0.688 a 0.825

<sup>a</sup> DeLong et al., 1988

<sup>b</sup> Binomial exact

#### Comparación de las curvas ROC por pares

<b>BDI-II ~ Puntuación directa de la escala Distimia</b>	
Diferencia entre áreas	0.00599
Error Estándar <sup>c</sup>	0.0298
Intervalo de confianza al 95%	-0.0524 a 0.0644
Estadístico z	0.201
Nivel de significación	P = 0.8408
<b>BDI-II ~ Puntuación directa de la escala Depresión Mayor</b>	
Diferencia entre áreas	0.0213
Error Estándar <sup>c</sup>	0.0392
Intervalo de confianza al 95%	-0.0556 a 0.0983
Estadístico z	0.543
Nivel de significación	P = 0.5869
<b>BDI- I ~ Puntuación tasa base de la escala Distimia</b>	
Diferencia entre áreas	0.0127
Error Estándar <sup>c</sup>	0.0298
Intervalo de confianza al 95%	-0.0457 a 0.0711
Estadístico z	0.426
Nivel de significación	P = 0.6698
<b>BDI-II ~ Puntuación tasa base de la de la escala Depresión Mayor</b>	
Diferencia entre áreas	0.0228
Error Estándar <sup>c</sup>	0.0421
Intervalo de confianza al 95%	-0.0598 a 0.105

Estadístico z	0.541
Nivel de significación	P = 0.5884
Puntuación directa de la escala Distimia ~ Puntuación directa de la escala Depresión Mayor	
Diferencia entre áreas	0.0273
Error Estándar <sup>c</sup>	0.0304
Intervalo de confianza al 95%	-0.0323 a 0.0869
Estadístico z	0.898
Nivel de significación	P = 0.3693
Puntuación directa de la escala Distimia ~ Puntuación tasa base de la escala Distimia	
Diferencia entre áreas	0.00672
Error Estándar <sup>c</sup>	0.0106
Intervalo de confianza al 95%	-0.0140 a 0.0275
Estadístico z	0.635
Nivel de significación	P = 0.5256
Puntuación directa de la escala Distimia ~ Puntuación tasa base de la escala Depresión Mayor	
Diferencia entre áreas	0.0288
Error Estándar <sup>c</sup>	0.0337
Intervalo de confianza al 95%	-0.0372 a 0.0948
Estadístico z	0.855
Nivel de significación	P = 0.3928
Puntuación directa de la escala Depresión Mayor ~ Puntuación tasa base de la escala Distimia	
Diferencia entre áreas	0.0340
Error Estándar <sup>c</sup>	0.0323
Intervalo de confianza al 95%	-0.0293 a 0.0974
Estadístico z	1.053
Nivel de significación	P = 0.2924
Puntuación directa de la escala Depresión Mayor ~ Puntuación tasa base de la escala Depresión Mayor	
Diferencia entre áreas	0.00147
Error Estándar <sup>c</sup>	0.0136
Intervalo de confianza al 95%	-0.0251 a 0.0280
Estadístico z	0.108
Nivel de significación	P = 0.9136
Puntuación tasa base de la escala Distimia ~ Puntuación tasa base de la escala Depresión Mayor	
Diferencia entre áreas	0.0355
Error Estándar <sup>c</sup>	0.0347
Intervalo de confianza al 95%	-0.0324 a 0.103
Estadístico z	1.024
Nivel de significación	P = 0.3057

<sup>c</sup> DeLong et al., 1988

**Anexo 15.** Comparación de curvas ROC de las puntuaciones directas y tasa base de las escalas de Distimia y Depresión Mayor del MCMI-III con respecto al diagnóstico de trastorno depresivo mayor asignado en la SCID-I

Variable 1	BDI-II
Variable 2	Puntuación directa de la escala Depresión Mayor
Variable 3	Puntuación tasa base de la escala Depresión Mayor
Variable 4	Puntuación directa de la escala Distimia
Variable 5	Puntuación tasa base de la escala Distimia
Variable de clasificación	Diagnóstico de trastorno depresivo mayor (asignado en la SCID-I)

	AUC	SE <sup>a</sup>	95% CI <sup>b</sup>
BDI- II	0.860	0.0272	0.800 a 0.908
Puntuación directa de la escala de Depresión Mayor	0.842	0.0292	0.779 a 0.892
Puntuación tasa base de la escala Depresión Mayor	0.838	0.0293	0.776 a 0.889
Puntuación directa de la escala Distimia	0.825	0.0307	0.761 a 0.878
Puntuación tasa base de la escala Distimia	0.822	0.0312	0.758 a 0.875

<sup>a</sup> DeLong et al., 1988

<sup>b</sup> Binomial exact

#### Comparación de las curvas ROC por pares

<b>BDI-II ~ Puntuación directa de la escala Depresión Mayor</b>	
Diferencia entre áreas	0.0186
Error Estándar <sup>c</sup>	0.0326
Intervalo de confianza al 95%	-0.0452 a 0.0824
Estadístico z	0.571
Nivel de significación	P = 0.5677
<b>BDI-II ~ Puntuación directa de la escala Distimia</b>	
Diferencia entre áreas	0.0354
Error Estándar <sup>c</sup>	0.0314
Intervalo de confianza al 95%	-0.0262 a 0.0970
Estadístico z	1.125
Nivel de significación	P = 0.2605
<b>BDI-II ~ Puntuación tasa base de la escala Depresión Mayor</b>	
Diferencia entre áreas	0.0220
Error Estándar <sup>c</sup>	0.0345
Intervalo de confianza al 95%	-0.0455 a 0.0896
Estadístico z	0.640
Nivel de significación	P = 0.5222
<b>BDI-II ~ Puntuación tasa base de la escala Distimia</b>	
Diferencia entre áreas	0.0383
Error Estándar <sup>c</sup>	0.0321
Intervalo de confianza al 95%	-0.0246 a 0.101
Estadístico z	1.194

Nivel de significación	P = 0.2325
Puntuación directa de la escala Depresión Mayor ~ Puntuación directa de la escala Distimia	
Diferencia entre áreas	0.0168
Error Estándar <sup>c</sup>	0.0251
Intervalo de confianza al 95%	-0.0324 a 0.0659
Estadístico z	0.669
Nivel de significación	P = 0.5037
Puntuación directa de la escala Depresión Mayor ~ Puntuación tasa base de la escala Depresión Mayor	
Diferencia entre áreas	0.00343
Error Estándar <sup>c</sup>	0.0119
Intervalo de confianza al 95%	-0.0200 a 0.0268
Estadístico z	0.287
Nivel de significación	P = 0.7737
Puntuación directa de la escala Depresión Mayor ~ Puntuación tasa base de la escala Distimia	
Diferencia entre áreas	0.0197
Error Estándar <sup>c</sup>	0.0275
Intervalo de confianza al 95%	-0.0341 a 0.0736
Estadístico z	0.717
Nivel de significación	P = 0.4731
Puntuación directa de la escala Distimia ~ Puntuación tasa base de la escala Depresión Mayor	
Diferencia entre áreas	0.0133
Error Estándar <sup>c</sup>	0.0256
Intervalo de confianza al 95%	-0.0369 a 0.0635
Estadístico z	0.520
Nivel de significación	P = 0.6028
Puntuación directa de la escala Distimia ~ Puntuación tasa base de la escala Distimia	
Diferencia entre áreas	0.00295
Error Estándar <sup>c</sup>	0.0103
Intervalo de confianza al 95%	-0.0173 a 0.0232
Estadístico z	0.286
Nivel de significación	P = 0.7752
Puntuación tasa base de la escala Depresión Mayor ~ Puntuación tasa base de la escala Distimia	
Diferencia entre áreas	0.0163
Error Estándar <sup>c</sup>	0.0274
Intervalo de confianza al 95%	-0.0374 a 0.0700
Estadístico z	0.594
Nivel de significación	P = 0.5526

<sup>c</sup> DeLong et al., 1988

# **REFERENCIAS BIBLIOGRÁFICAS**



## Referencias bibliográficas

- Aboraya, A., France, C., Young, J., Curci, K., y LePage, J. (2005). The validity of psychiatric diagnosis revisited: the clinician's guide to improve the validity of psychiatric diagnosis. *Psychiatry*, 2(9), 48-55.
- Alarcón, R. D (2010). Hacia nuevos sistemas de diagnóstico: proceso, preguntas y dilemas. *Revista de Psiquiatría y Salud Mental*, 3(2), 37-39.
- Alegría, M., Shrout, P. E., Torres, M., Lewis-Fernández, R., Abelson, J. M., Powell, M. Interian, A., Lin, J., Laderman, M. y Canino, G. (2009). Lessons learned from the clinical reappraisal study of the Composite International Diagnostic Interview with latinos. *International Journal of Methods in Psychiatric Research*, 18(2), 84-95.
- Alonso, J., Angermeyer, M. C., Bernert, S., Bruffaerts, R., Brugha, T. S., Bryson, H., de Girolamo, G., Graaf, R., Demyttenaere, K., Gasquet, I., Haro, J. M., Katz, S. J., Kessler, R. C., Kovess, V., Lépine, J. P., Ormel, J., Polidori, G., Russo, L. J., Vilagut, G., Almansa, J., Arbabzadeh-Bouchez, S., Autonell, J., Bernal, M., Buist-Bouwman, M. A., Codony, M., Domingo-Salvany, A., Ferrer, M., Joo, S. S., Martínez-Alonso, M., Matschinger, H., Mazzi, F., Morgan, Z., Morosini, P., Palacín, C., Romera, B., Taub, N. y Vollebergh, W. A. (2004a). 12-month comorbidity patterns and associated factors in Europe: Results from the European Study of the Epidemiology of Mental Disorders (ESEMeD) project. *Acta Psychiatrica Scandinavica, June Supplement*, 109(420), 28-37.
- Alonso, J., Angermeyer, M. C., Bernert, S., Bruffaerts, R., Brugha, T. S., Bryson, H., de Girolamo, G., Graaf, R., Demyttenaere, K., Gasquet, I., Haro, J.M., Katz, S. J., Kessler, R. C., Kovess, V., Lépine, J. P., Ormel, J., Polidori, G., Russo, L. J., Vilagut, G., Almansa, J., Arbabzadeh-Bouchez, S., Autonell, J., Bernal, M., Buist-Bouwman, M. A., Codony, M., Domingo-Salvany, A., Ferrer, M., Joo, S. S., Martínez-Alonso, M., Matschinger, H., Mazzi, F., Morgan, Z., Morosini, P., Palacín, C., Romera, B., Taub, N. y Vollebergh, W. A. (2004b). Disability and quality of life impact of mental disorders in Europe: Results from the European Study of the Epidemiology of Mental Disorders (ESEMeD) project. *Acta Psychiatrica Scandinavica, June Supplement*, 109(420), 38-46.
- Altman, D. G. (1991). *Practical statistics for medical research*. New York: Chapman and Hall.



- American Psychiatric Association (1983). *DSM-III: Manual Diagnóstico y Estadístico de los Trastornos Mentales*. Barcelona: Masson.
- American Psychiatric Association (2002). *DSM-IV-TR. Manual Diagnóstico y Estadístico de los Trastornos Mentales. Texto revisado*. Barcelona: Masson.
- American Psychiatric Association (2012). *Major depressive disorder, single episode. DSM-5 development*. Recuperado el 13 de diciembre de 2012. <http://www.dsm5.org/Lists/ProposedRevision/DispForm.aspx?ID=44>.
- Andrews, G., Brugha, T., Thase, M. E., Duffy, F. F., Rucci, P., y Slade, T. (2007). Dimensionality and the category of major depressive episode. En J. E. Helzer, H. C. Kraemer, R. F. Krueger, H. Wittchen, P. J. Sirovatka y D. A. Regier (Eds.), *Dimensional approaches in diagnostic classification. Refining the research agenda for DSM-V* (pp. 35-51). Washington, DC: American Psychiatric Association.
- Andrews, G., y Peters, L. (1998). The psychometric properties of the Composite International Diagnostic Interview. *Social Psychiatry and Psychiatric Epidemiology*, 33(2), 80-88.
- Andrews, G., Peters, L., Guzman, A. y Bird, K. (1995). A comparison of two structured diagnostic interviews: CIDI and SCAN. *Australian and New Zealand Journal of Psychiatry*, 29(1), 124-132.
- Aragonès Benaiges, E., Masdéu Montalà, R. M., Cando Guasch, G., y Coll Borràs, G. (2001). Validez diagnóstica de la Self-rating Depression Scale de Zung en pacientes de atención primaria. *Actas Españolas De Psiquiatría*, 29(5), 310-316.
- Arroll, B., Goodyear-Smith, F., Crengle, S., Gunn, J., Kerse, N., Fishman, T., Falloon, K. y Hatcher, S. (2010). Validation of PHQ-2 and PHQ-9 to screen for major depression in the primary care population. *Annals of Family Medicine*, 8(4), 348-353.
- Ayuso-Mateos, J.L., Gutiérrez Recacha, P., Haro, J.M., Salvador Carulla, L., Vázquez Polo, F.J., Negrín Hernández, M.A., y Chisholm, D. (2008). *Reducing the burden of mental illness in Spain. Population-level impact and cost-effectiveness of treatments in depression and schizophrenia*. Bilbao: Fundación BBVA.
- Ayuso-Mateos, J.L., Vázquez-Barquero, J.L., Dowrick, C., Lehtinen, V., Dalgard, O.S., Casey, P., Wilkinson, C., Lasa, L., Page, H., Dunn, G. y Wilkinson, G. (2001).

- Depressive disorders in Europe: prevalence figures from the ODIN study. *British Journal of Psychiatry*, 179(4), 308-316.
- Baca, E., Saiz, J., Agüera, L., Caballero, L., Fernández-Liria, A., Ramos, J., Gil, A., Madrigal, M. y Porras, A. (1999). Validación de la versión española del PRIME-MD: un procedimiento para el diagnóstico de trastornos mentales en atención primaria. *Actas Españolas De Psiquiatría*, 27(6), 375-383.
- Bagby, R. M., Ryder, A. G., Schuller, D. R., y Marshall, M. B. (2004). The Hamilton Depression Rating Scale: Has the gold standard become a lead weight? *The American Journal of Psychiatry*, 161(12), 2163-2177.
- Bakker, I. M., Terluin, B., van Marwijk, Harm W. J., van Mechelen, W., y Stalman, W. A. B. (2009). Test-retest reliability of the PRIME-MD: Limitations in diagnosing mental disorders in primary care. *European Journal of Public Health*, 19(3), 303-307.
- Ballesteros, J., Bobes, J., Bulbena, A., Luque, A., Dal-Ré, R., Ibarra, N., y Güemes, I. (2007). Sensitivity to change, discriminative performance, and cutoff criteria to define remission for embedded short scales of the Hamilton Depression Rating Scale (HAM-D). *Journal of Affective Disorders*, 102(1-3), 93-99.
- Basco, M. R., Bostic, J. Q., Davies, D., Rush, A. J., Witte, B., Hendrickse, W., y Barnett, V. (2000). Methods to improve diagnostic accuracy in a community mental health setting. *American Journal of Psychiatry*, 157, 1599-1605.
- Bech, P. (1996). *The Bech, Hamilton and Zung scales for mood disorders: screening and listening. A twenty years update with reference to DSM-IV and ICD-10*. 2ª ed.. Berlin: Springer-Verlag.
- Beck, A. T. y Alford, B. A. (Eds.) (2009). The definition of depression. *Depression: causes and treatments*, 2ª ed. (pp. 3-11). Philadelphia, PA: University of Pennsylvania Press.
- Beck, A. T., Rush A. J., Shaw, B. F. y Emery, G. (2005). *Terapia cognitiva para la depresión*, 16ª Ed.. Bilbao: Desclée de Brouwer.
- Beck, A. T., y Steer, R. A. (1993). *Beck Depression Inventory. Manual*. San Antonio, Texas: The Psychological Corporation.

- Beck, A. T., Steer, R. A. y Brown, G. K. (1996). *BDI-II. Beck Depression Inventory-Second Edition. Manual*. San Antonio, Texas: The Psychological Corporation.
- Beck, A. T., Steer, R. A., y Brown, G. K. (2000). *BDI-Fast Screen for medical patients, Manual*. San Antonio, Texas: The Psychological Corporation.
- Beck, A. T., Steer, R. A., y Brown, G. K. (2011a). *Manual. BDI-II. Inventario de Depresión de Beck-II* (Adaptación española: Sanz, J., y Vázquez, C.). Madrid: Pearson Educación.
- Beck, A. T., Steer, R. A., y Brown, G. K. (2011b). *Manual. BDI-FastScreen. Inventario de Depresión de Beck FastScreen para pacientes médicos* (Adaptación española: Sanz, J., Izquierdo, A., y García-Vera, M. P.). Madrid: Pearson Educación.
- Beck, A. T., Ward, C. H., Mendelson, M., Mock, J. y Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, 4, 561-571.
- Bentall, R. P. (2006). Diagnosis of schizophrenia: a 19th Century concept not fit for 21st Century science. *Future Neurology*, 1(3), 241-244.
- Berrios, G. E. (2008). *Historia de los síntomas de los trastornos mentales. La psicopatología descriptiva desde el siglo XIX*. México: Fondo de Cultura Económica.
- Bethencourt, J. M., Peñate, W., González, M. y Fumero, A. (1996). El programa de entrevista diagnóstica estructurada (D.I.S.): presentación de la versión española y primeros datos de validez a través de la fiabilidad intercalificadores. *Psicologemas*, 10, 47-154.
- Bijl, R. V., Ravelli, A., y van Zessen, G. (1998). Prevalence of psychiatric disorder in the general population: Results of the Netherlands Mental Health Survey and Incidence Study (NEMESIS). *Social Psychiatry and Psychiatric Epidemiology*, 33(12), 587-595.
- Biometrics Research Department (n.d.). *What is the Reliability of the SCID-I?* Recuperado el 21 de febrero de 2012. [http://www.cumc.columbia.edu/dept/scid/psychometric/scidI\\_reliability.html](http://www.cumc.columbia.edu/dept/scid/psychometric/scidI_reliability.html).

- Blashfield, R. K. y Livesley W. J. (2008). Classification. En T. Millon, P. H. Blaney, R. D. Davis (Eds.), *Oxford textbook of psychopathology* (pp. 3-28). New York: Oxford University Press.
- Bobes, J. (1998). A Spanish validation study of the mini international neuropsychiatric interview. *European Psychiatry*, 13(Supp. 4), 198-199.
- Bobes, J., Bulbena, A., Luque, A., Dal-Re, R., Ballesteros, J. e Ibarra, N. (2003). A comparative psychometric study of the Spanish versions with 6, 17, and 21 items of the Hamilton Depression Rating Scale. *Medicina Clínica*, 120(18), 693-700.
- Bobes, J., Portilla, M. P. G., Bascarán, M. T., Saiz, P. A. y Bousoño, M. (2004). Instrumentos de evaluación para los trastornos del humor. En *Banco de instrumentos básicos para la práctica de la psiquiatría clínica*, 3ª ed. (pp. 51-69). Barcelona: Ars Médica.
- Bonanno, G. A., Goorin, L., y Coifman, K. G. (2008). Sadness and grief. M. Lewis, J. Haviland-Jones y L. Feldman Barrett (Eds.), *The handbook of emotion (3rd Edition)*. New York: Guilford.
- Bostwick, J. M., y Pankratz, V. S. (2000). Affective disorders and suicide risk: A reexamination. *The American Journal of Psychiatry*, 157(12), 1925-1932.
- Bromet, E., Andrade, L. H., Hwang, I., Sampson, N. A., Alonso, J., Girolamo, G., Graaf, R., Demyttenaere, K., Hu, C., Iwata, N., Karam, A. N., Kaur, J., Kostyuchenko, S., Lépine, J. P., Levinson, D., Matschinger, H., Medina, M. H., Browne, M. O., Posada-Villa, J., Viana, M. C., Williams, D. R. y Kessler R. C. (2011). Cross-national epidemiology of DSM-IV major depressive episode. *BioMed Central Medicine*, 9(1), 90-106.
- Brophy, C.J., Norvell N. K. y Kiluk, D.J. (1988). An examination of the factor structure and convergent and discriminant validity of the SCL-90R in an outpatient clinic population. *Journal of Personality*, 52(2), 334-340.
- Brugha, T. S., Jenkins, R., Taub, N., Meltzer, H., y Bebbington, P. E. (2001). A general population comparison of the Composite International Diagnostic Interview (CIDI) and the Schedules for Clinical Assessment in Neuropsychiatry (SCAN). *Psychological Medicine*, 31(6), 1001-1013.

- Bucholz, K. K., Marion, S. L., Shayka, J. J., Marcus, S. C. y Robins, L. N. (1996). A short computer interview for obtaining psychiatric diagnoses. *Psychiatric Services*, 47(3), 293-297.
- Bufka, L. F., y Camp, N. (2010). *Brief measures for screening and measuring mental health outcomes*. En Antony M. M., Barlow D. H. (Eds.). *Handbook of assessment and treatment planning for psychological disorders*, (pp. 62-94). New York: Guilford Press.
- Bourdon, K. H., Rae, D. S., Locke, B. Z., Narrow, W. E., y Regier, D. A. (1992). Estimating the prevalence of mental disorders in U.S. adults from the Epidemiologic Catchment Area Survey. *Public Health Reports*, 107(6), 663–668.
- Burton, R. (2008). *Anatomía de la melancolía*. Madrid: Alianza Editorial. (Original publicado en 1621).
- Cano Vindel, A., Martín Salguero, J., Mae Wood, C., Dongil, E., y Latorre, J. M. (2012). La depresión en atención primaria: prevalencia, diagnóstico y tratamiento. *Papeles del Psicólogo*, 33(1), 2-11.
- Cardenal V., Sánchez, M. P., y Ortiz-Tallo M. (2007a). Los trastornos de personalidad según el modelo de Millon: una propuesta integradora. *Clínica y Salud*, 18(3), 305-324.
- Cardenal V., Sánchez, M. P., y Ortiz-Tallo M. (2007b). *Guía práctica para la interpretación del MCMI-III: evaluación de 14 casos clínicos*. Madrid: TEA Ediciones.
- Carnero-Pardo, C. (2005). La evaluación de las pruebas diagnósticas. *Revista de Neurología*, 40, 641-643.
- Cavanagh, J. T. O., Carson, A. J., Sharpe, M., y Lawrie, S. M. (2003). Psychological autopsy studies of suicide: A systematic review. *Psychological Medicine*, 33(3), 395-405.
- Choca, J., Bresolin, L., Okonek, A., y Ostrow, D. (1988). Validity of the Millon Clinical Multiaxial Inventory in the assessment of affective disorders. *Journal of Personality Assessment*, 52, 96-105.
- Clark, D. A., Beck, A. T. y Alford, B. A. (1999). Depressive Symptoms and Disorders: Diagnostic and Conceptual Considerations. En D. A. Clark, A. T. Beck, B. A.

- Alford (Eds.), *Scientific foundations of cognitive theory and therapy of depression* (pp. 1-35). New York: John Wiley and Sons.
- Cochrane, A. L. (1972). *Effectiveness and efficiency: random reflection on health services*. London: Nuffield Provincial Hospitals Trust.
- Comisión Internacional de Tests (ITC) (2000). Directriz 2.7.9. *Directrices Internacionales para el uso de los Tests*. Recuperado el 20 de febrero de 2012. <http://www.cop.es/index.php?page=directrices-internacionales>.
- Conde, V., Escribá, J. A., e Izquierdo, J. A. (1970). Evaluación estadística y adaptación castellana de la Escala Autoaplicada para la Depresión de Zung. *Archivos de Neurobiología. Parte I*, XXXIII, 2, 185-206, y *Parte II*, XXXIII, 3, 201-203.
- Conde, V., y Franch, J. L. (1984). *Escalas de evaluación comportamental para la cuantificación de la sintomatología en los trastornos angustiosos y depresivos*. Madrid: Upjohn Farmoquímica.
- Conti, N. A. (2007). *Historia de la depresión. La melancolía desde la Antigüedad hasta el Siglo XIX*. Buenos Aires: Editorial Polemos.
- Coryell, W. y Young, E. A. (2005). Clinical predictors of suicide in primary major depressive disorder. *Journal of Clinical Psychiatry*, 66(4), 412-17.
- Coryell, W., Winkur, G., Maser, J.D., Akiskal, H. S., Keller, M. B. y Edicott, J. (1994). Recurrently situational (reactive) depression: A study of course, phenomenology and familial psychopathology. *Journal of affective disorders*, 31, 203-210.
- Craig, R. J. (1999). Overview and current status of the millon clinical multiaxial inventory. *Journal of Personality Assessment*, 72(3), 390-406.
- Craig, R. J. (2005a). Alternative interpretations for the Histrionic, Narcissistic and Compulsive personality disorder scales of the MCMI-III. En R. J. Craig (Ed.), *New directions in interpreting the Millon Clinical Multiaxial Inventory-III* (MCMI-III) (pp. 71-93). Hoboken, New Jersey: John Wiley and Sons.
- Craig, R. J. (Ed.). (2005b). *New directions in interpreting the Millon Clinical Inventory-III (MCMI-III)*. Hoboken, New Jersey: John Wiley and Sons.
- Colman, I. y Ataullahjan, A. (2010). Life course perspectives on the epidemiology of depression. *Canadian Journal of Psychiatry*, 55(10), 622-632.

- Cuijpers, P. y Smit, F. (2004). Subthreshold depression as a risk indicator for major depressive disorder: A systematic review of prospective studies. *Acta Psychiatrica Scandinavica*, 109(5), 325-331.
- Cuijpers, P., Smit, F. y Willemse, G. (2005). Predicting the onset of major depression in subjects with subthreshold depression in primary care: A prospective study. *Acta Psychiatrica Scandinavica*, 111, 133-138.
- Demyttenaere, K., Bruffaerts, R., Posada-Villa, J., Gasquet, I., Kovess, V., Lepine, J. P., Angermeyer, M. C., Bernert, S., de Girolamo, G., Morosini, P., Polidori, G., Kikkawa, T., Kawakami, N., Ono, Y., Takeshima, T., Uda, H., Karam, E. G., Fayyad, J. A., Karam, A. N., Mneimneh, Z. N., Medina-Mora, M. E., Borges, G., Lara, C., de Graaf, R., Ormel, J., Gureje, O., Shen, Y., Huang, Y., Zhang, M., Alonso, J., Haro, J. M., Vilagut, G., Bromet, E. J., Gluzman, S., Webb, C., Kessler, R. C., Merikangas, K. R., Anthony, J. C., Von Korff, M. R., Wang, P. S., Brugha, T. S., Aguilar-Gaxiola, S., Lee, S., Heeringa, S., Pennell, B. E., Zaslavsky, A. M., Üstün, T. B. y Chatterji, S. (2004). Prevalence, severity, and unmet need for treatment of mental disorders in the World Health Organization World Mental Health Surveys. *JAMA: Journal of American Medicine Association*, 291(21), 2581-2590.
- Derogatis, L. R. (1977) *SCL-90-R. Administration, Scoring and Procedures Manual I for the Revised Version of the SCL-90*. Baltimore: John Hopkins University Press.
- Derogatis, L. R. (2002). *SCL-90-R. Cuestionario de 90 Síntomas-Revisado* (Adaptación española: González de Rivera, J. L., de las Cuevas, C., Rodríguez Abuín, M. y Rodríguez Pulido, F.). Madrid: TEA Ediciones.
- Diez-Quevedo, C., Rangil, T., Sanchez-Planell, L., Kroenke, K y Spitzer, R. L. (2001). Validation and utility of the patient health questionnaire in diagnosing mental disorders in 1003 general hospital Spanish inpatients. *Psychosomatic Medicine*, 63(4), 679-686.
- Dimidjian, S., Barrera, M., Martell, C., Muñoz, R. F., y Lewinsohn, P. M. (2011). The origins and current status of behavioral activation treatments for depression. *Annual Review of Clinical Psychology*, 7, 1-38.
- Dozois, D. J. A. y Dobson, K. S. (2011). Depression. En M. M. Antony y D. H. Barlow (Eds.), *Handbook of assessment and treatment planning for psychological disorders*, 2ª ed. (pp. 344-389). New York: Guilford Press.

- Dozois, D. J. A., Dobson, K., S y Ahnberg, J. L. (1998). A Psychometric Evaluation of the Beck Depression Inventory-II. *Psychological Assessment*, 10(2), 83-89.
- Eaton, W.W., Neufeld, K., Chen, L., y Cai, G. (2000). A comparison of self-report and clinical diagnostic interviews for depression: Diagnostic Interview Schedule and Schedules for Clinical Assessment in Neuropsychiatry in the Baltimore Epidemiologic Catchment Area follow-up. *Archives of General Psychiatry*, 57(3), 217-222.
- Eaton, N. R., South, S. C. y Krueger, R. F. (2010). The Meaning of comorbidity among common mental disorders. En T. Millon, R. F. Frueger y E. Simonsen (Eds.), *Contemporary directions in psychopathology: scientific foundations of the DSM-V and ICD-10*, (pp. 223-241). New York: Guilford Press.
- Ebmeier, K. P., Donaghey, C., y Steele, J. D. (2006). Recent developments and current controversies in depression. *Lancet*, 367(9505), 153-167.
- Epidemiology and Prevention Research Group (2010). *Description of the C DIS-IV*. Washington University School of Medicine in St. Louis. Recuperado el 21 de febrero de 2012: <http://epi.wustl.edu/CDISIV/Description.aspx>.
- Estupiñá, F., Labrador, F. J., y García-Vera, M. P. (2012). A study of patients who go to a psychology clinic seeking treatment. *Spanish Journal of Psychology*, 15(1), 275-285.
- Evers, A., Muñiz, J., Bartram, D., Boben, D., Egeland, J., Fernández-Hermida, J. R., Frans, O., Gintiliené, G., Hagemester, C., Halama, P., Iliescu, D., Jaworowska, D., Jiménez, P., Manthouli, M., Matesic, K., Schittekatte, M., Sümer, H. C., y Urbánek, T. (2012). Testing practices in the 21th century. Developments and European psychologists' opinions. *European Psychologists*, 17(14), 300-319.
- Fan, J., Upadhye, S. y Worster, A. (2006). Understanding receiver operating characteristic (ROC) curves. *Canadian Journal of Emergency Medicine*, 8(1), 19-20.
- Fava, M., Hwang, I., Rush, A. J., Sampson, N., Walters, E. E., y Kessler, R. C. (2010). The importance of irritability as a symptom of major depressive disorder: Results from the National Comorbidity Survey Replication. *Molecular Psychiatry*, 15(8), 856-867.



- Fawcett, T. (2004). ROC graphs: notes and practical considerations for researchers. *Kluwer Academic Publishers*. Recuperado el 21 de febrero de 2012. [http://www.hpl.hp.com/techreports/2003/HPL-2003-4.pdf?jumpid=reg\\_RI002JJSENGamma](http://www.hpl.hp.com/techreports/2003/HPL-2003-4.pdf?jumpid=reg_RI002JJSENGamma)
- Feliciano, L. y Gum A. M. (2010). Mood Disorders. En D. L. Segal y M. Hersen (Eds.), *Diagnostic interviewing*, (pp. 153-176). New York: Springer Publishing.
- Fennig, S., Craig, T., Lavelle, J., Kovasznay, B., y Bromet, E. J. (1994). Best-estimate versus structured interview-based diagnosis in first-admission psychosis. *Comprehensive Psychiatry*, 35(5), 341-348.
- Fennig, S., Naisberg-Fennig, S., Craig, T. J., Tanenberg-Karant, M., y Bromet, E. J. (1996). Comparison of clinical and research diagnoses of substance use disorders in a first-admission psychotic sample. *The American Journal on Addictions*, 5(1), 40-48.
- Fernández, A., Pinto-Meza, A., Bellón, J. A., Roura-Poch, P., Haro, J. M., Autonell, J., Palao, D. J., Peñarrubia, M. T., Fernández, R., Blanco, E., Luciano, J. V. y Serrano-Blanco, A. (2010). Is major depression adequately diagnosed and treated by general practitioners? Results from an epidemiological study. *General Hospital Psychiatry*, 32(2), 201-209.
- First, M. B., Pincus, H. A., Levine, J. B., Williams, J. B., Üstün, B., y Peele, R. (2004). Clinical utility as a criterion for revising psychiatric diagnoses. *The American Journal of Psychiatry*, 161(6), 946-954.
- First, M. B., Spitzer, R. L., Gibbon, M., y Williams, J. B. W. (1997). *Structured Clinical Interview for DSM-IV Axis I disorders (SCID-I)*. New York: Biometrics Research Department, New York Psychiatric Institute.
- First, M. B., Spitzer, R. L., Gibbon, M. y Williams, J. B. W. (1999). *Guía del usuario para la Entrevista Clínica Estructurada para los Trastornos del Eje I del DSM-IV. Versión clínica. SCID-I*. Barcelona: Masson.
- First, M. B., Spitzer, R. L., Gibbon, M. y Williams, J. B. W. (2002). *Structured Clinical Interview for DSM-IV Axis I Disorders, research version, non-patient edition (SCID-I/NP)*. New York: Biometrics Research Department, New York Psychiatric Institute.

- Furukawa, T. A., Watanabe, N y Omori, I. M. (2010). What (no) differences in responses to three classes of psychotropics can teach us about distinctions between generalized anxiety disorder and mayor depressive disorder. En D. P. Goldberg, K. S. Kendler y P. J. Sirovatka (Eds.), *Diagnostic issues in depression and generalized anxiety disorder: refining the research agenda for DSM-V* (pp. 71-104). Arlington, VA: American Psychiatric Association.
- Gabilondo, A., Rojas-Farreras, S., Vilagut, G., Haro, J. M., Fernández, A, Pinto-Meza, A y Alonso, A. (2010). Epidemiology of major depressive episode in a southern European country: Results from the ESEMeD-Spain project. *Journal of Affective Disorders*, 120(1–3), 76-85.
- Gaminde, I., Uria, M., Padro, D. y Querejeta, I. (1993). Depression in three populations in the Basque country –a comparison with Britain. *Social Psychiatry and Psychiatric Epidemiology*, 28(5), 243-251.
- Ganellen, R. J. (1996). Comparing the diagnostic efficiency of the MMPI, MCMI-II, and Rorschach: a review. *Journal of Personality Assessment*, 67, 219-243.
- Gracia, E., y Herrero, J. (2004a). Determinants of social integration in the community: an exploratory analysis of personal, interpersonal, and situational variables. *Journal of Community and Applied Social Psychology*, 14, 1-15.
- Gracia, E., y Herrero, J. (2004b). Personal and situational determinants of relationship-specific perceptions of social support. *Social Behavior and Personality*, 32, 459-476.
- García-Vera, M. P., y Sanz, J. (2009). La Clínica Universitaria de Psicología de la Universidad Complutense de Madrid: historia, características y situación actual. *Acción Psicológica*, 6, 123-134.
- García-Vera, M. P., y Sanz, J. (2012). Depresión. En M<sup>a</sup>. I. Comeche Moreno y M. A. Vallejo Pareja (Eds.), *Manual de terapia de conducta en la infancia*, 2<sup>a</sup> edición (pp. 197-241). Madrid: Dykinson.
- Gibeau, P., y Choca, J. (2005). The diagnostic efficiency of the MCMI-III in the detection of Axis I disorders. En R. J. Craig (Ed.), *New directions in interpreting the Millon Clinical Multiaxial Inventory-III (MCMI-III)* (pp. 272-283). Hoboken, NJ: John Wiley & Sons.

- Gilbody, S., Bower, P., Fletcher, J., Richards, D. y Sutton, J.D. (2006). Collaborative care for depression. A cumulative meta-analysis and review of longer-term outcomes. *Archives of International Medicine*, 166, 2314-2321.
- Gilbody S., Richards, D., Brealey S. y Hewitt, C. (2007). Screening for Depression in Medical Settings with the Patient Health Questionnaire (PHQ): A Diagnostic Meta-Analysis. *Journal of General Internal Medicine*, 22(11), 1596-1602.
- Gloster, A. T., Rhoades, H. M., Novy, D., Klotsche, J., Senior, A., Kunik, M., Wilson N. y Stanley, M.A. (2008). Psychometric properties of the depression anxiety and stress scale-21 in older primary care patients. *Journal of Affective Disorders*, 110(3), 248-259.
- Goldberg, D., y Huxley, P. (1980). *Mental illness in the community: the pathway to psychiatric care*. London: Tavistock Publications.
- Goldney, R. D. (2006). The Utility of the DSM Nosology of Mood Disorders. *The Canadian Journal of Psychiatry*, 51(14), 874-878.
- Goodwin, R. D., Jacobi, F., Bittner, A., Wittchen, H.-U., Stein, D. J., Kupfer, D. J., Schatzberg, A.F. (2006). Epidemiology of mood disorders. En D. J. Stein, D. J. Kupfer y A. F. Schatzrberg (Eds.), *The American Psychiatric Publishing textbook of mood disorders* (pp. 33-54). Arlington, VA: American Psychiatric Publishing.
- de Graaf, R., Ten Have, M., van Gool, C. y van Dorsselaer, S. (2011). Prevalence of mental disorders and trends from 1996 to 2009. Results from the Netherlands Mental Health Survey and Incidence Study-2. *Social psychiatry and psychiatric epidemiology*, 47(2), 203-213.
- Grove, W. M. y Vrieze S. I. (2009). An exploration of the base rate scores of the Millon Clinical Multiaxial Inventory–III. *Psychological Assessment*, 21, 57-67.
- Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery and Psychiatry*, 23, 56-62.
- Hanel, G., Henningsen, P., Herzog, W., Sauer, N., Schaefer, R., Szecsenyi, J., y Löwe, B. (2009). Depression, anxiety, and somatoform disorders: Vague or distinct categories in primary care? Results from a large cross-sectional study. *Journal of Psychosomatic Research*, 67(3), 189-197.

- Hardeveld, F., Spijker, J., De Graaf, R., Nolen, W. A. y Beekman, A. T. F. (2010). Prevalence and predictors of recurrence of major depressive disorder in the adult population. *Acta Psychiatrica Scandinavica*, 122(3), 184-191.
- Haro, J. M., Arbabzadeh-Bouchez, S., Brugha, T. S., de Girolamo, G., Guyer, M. E., Jin, R., Lepine, J-P., Mazzi, F., Reneses, B., Vilagut, G., Sampson, N. A. y Kessler, R. C. (2006b). Concordance of the Composite International Diagnostic Interview Version 3.0 (CIDI 3.0) with standardized clinical assessments in the WHO World Mental Health Surveys. *International Journal of Methods in Psychiatric Research*, 15(4), 167-180.
- Haro, J. M., Palacín, C., Vilagut, G., Martínez, M., Bernal, M., Luque, I., Codony, M., Dolz, M., Alonso, J. (2006a). Prevalencia de los trastornos mentales y factores asociados: resultados del estudio ESEMeD-España. *Medicina Médica*, 126(12), 445-451.
- Harvard School of Medicine (2007). National Comorbidity Survey (NCS) and National Comorbidity Survey Replication (NCS-R). Recuperado el 20 de febrero de 2012 en <http://www.hcp.med.harvard.edu/ncs/>.
- Hasin, D. S., Goodwin, R. D., Stinson, F. S., y Grant, B. F. (2005). Epidemiology of major depressive disorder: Results from the national epidemiologic survey on alcoholism and related conditions. *Archives of General Psychiatry*, 62(10), 1097-1106.
- Hawton, K. y von Heeringen K. (2009). Suicide. *Lancet*, 373(9672), 1372-1378.
- Haynes, B. (1999). Can it work? Does it work? Is it worth it? The testing of healthcare interventions is evolving. *British Medical Journal*, 319, 652-653.
- Herrero, J. y Gracia, E. (2007). Una medida breve de la sintomatología depresiva (CESD-7). *Salud Mental*, 30(5), 40-46.
- Hettema, J. M. (2010). The nosologic relationship between generalized anxiety disorder and mayor depression. En D. P. Goldberg, K. S. Kendler y P. J. Sirovatka (Eds.), *Diagnostic issues in depression and generalized anxiety disorder: refining the research agenda for DSM-V* (pp. 15-40). Arlington, VA: American Psychiatric Association.

- Hirshbein, L. D. (2009). *American Melancholy: constructions of depression in the twentieth century*. Rutgers University Press.
- Horwitz, A. V. y Wakefield, J. C. (2007). *The loss of sadness. How psychiatry transformed normal sorrow into depressive disorder*. Oxford: Oxford University Press.
- Hunsley, J. y Mash, E. J. (2011). The role of assessment in evidence-based practice. En M. M. Antony, y D. H. Barlow (Eds.), *Handbook of assessment and treatment planning for psychological disorders* (pp. 3-22). New York: Guilford Press.
- Hsu, L. S. (2002). Diagnostic validity statistics and the MCMI-III. *Psychological Assessment*, 14, 410-422.
- Hsu, L. S. (2005). Using critiques of the MCMI to improve MCMI research and interpretations. En Craig, R. J. (Ed.), *New directions in interpreting the Millon Clinical Inventory-III (MCMI-III)* (pp. 165-184). Hoboken, New Jersey: John Wiley and Sons.
- Inch R. y Crossley M. (1993). Diagnostic utility of the MCMI-I and MCMI-II with psychiatric outpatients. *Journal of Clinical Psychology*, 49(3), 358-66.
- Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., Sanislow, C. y Wang, P. (2010). Research Domain Criteria (RDoC): Toward a new classification framework for research on mental disorders. *American Journal of Psychiatry*, 167(7), 748-751.
- Izquierdo Elizo, A. (2011). *Validación de la versión española del BDI-FS: un instrumento para la detección de la depresión en atención primaria*. Tesis doctoral. Facultad de Psicología, Universidad Complutense de Madrid. Consultada el 22 de diciembre de 2012 en: <http://eprints.ucm.es/14539/>
- Jablensky, A., y Kendell, R. E. (2002). Criteria for assessing a classification in Psychiatry. En M. Maj, W. Gaebel, J. J. López-Ibor y N. Sartorius (Eds.), *Psychiatric diagnosis and classification* (pp. 1-24). Chichester, UK: John Wiley & Sons.
- Jackson, S. W. (1989). *Historia de la melancolía y la depresión. Desde los tiempos hipocráticos a la época moderna*. Madrid: Ediciones Turner.

- Jacobi, F., Wittchen, H.-U., Holting, C., Hofler, M., Pfister, H., Müller, N., y Lieb, R. (2004). Prevalence, co-morbidity and correlates of mental disorders in the general population: Results from the German Health Interview and Examination Survey (GHS). *Psychological Medicine*, 34(4), 597-611.
- Jenkins, R., Bebbington, P., Brugha, T., Farrell, M., Gill, B. y Meltz, E. R. (1997). The National Psychiatric Morbidity Surveys of Great Britain —initial findings from the Household Survey. *Psychological Medicine*, 27(4), 775-789.
- Jenkins, R., Lewis, G., Bebbington, P., Brugha, T., Farrel, M., Gill, B. y Meltzer, H. (2003). The National Psychiatric Morbidity Surveys of Great Britain—initial findings from the Household Survey. *International Review of Psychiatry*, 15(1–2), 29–42.
- Joiner, T. E. (2010). *Myths about suicide*. Cambridge, MA: Harvard University Press.
- Joiner, T. E., Walker, R. L., Pettit, J. W., Perez, M., y Cukrowicz, K. C. (2005). Evidence-based assessment of depression in adults. *Psychological Assessment*, 17(3), 267-277.
- Jonas, S., Bebbington, P., McManus, S., Meltzer, H., Jenkins, R., Kuipers, E., Cooper, C., King, M. y Brugha, T. (2011). Sexual abuse and psychiatric disorder in England: Results from the 2007 adult psychiatric morbidity survey. *Psychological Medicine*, 41(4), 709-719.
- Jordanova, V., Wickramesinghe, C., Gerada, C., y Prince, M. (2004). Validation of two survey diagnostic interviews among primary care attendees: A comparison of CIS-R and CIDI with SCAN ICD-10 diagnostic categories. *Psychological Medicine*, 34(6), 1013-1024.
- Karlsson, L., Marttunen, M., Karlsson, H., Kaprio, J. y Hillevi, A. (2010). Minor change in the diagnostic threshold leads into major alteration in the prevalence estimate of depression. *Journal of Affective Disorders*, 122(1-2), 96–101.
- Kendell, R., y Jablensky, A. (2003). Distinguishing between the validity and utility of psychiatric diagnoses. *American Journal of Psychiatry*, 160(1), 4-12.
- Kendler, K. S. y Zachar, P. (2008). The incredible insecurity of psychiatric nosology. En K. S. Von Kendler y J. Parnas (Eds.), *Philosophical issues in psychiatry*:

- explanation, phenomenology, and nosology* (pp. 368-386). Baltimore: The Johns Hopkins University Press.
- Kessing, L. V. (2007). Epidemiology of subtypes of depression. *Acta Psychiatrica Scandinavica*, 115(Suppl. 433), 85-89.
- Kessler, R. C. (2007). The global burden of anxiety and mood disorders: putting the European Study of the Epidemiology of Mental Disorders (ESEMeD) findings into perspective. *Journal of Clinical Psychiatry*, 68(Suppl. 2), 10-19.
- Kessler, R. C., Abelson, J., Demler, O., Escobar, J. I., Gibbon, M., Guyer, M. E., Howes, M. J., Jin, R., Vega, W. A., Walters, E. E., Wang, P., Zaslavsky, A. y Zheng, H. (2004). Clinical calibration of DSM-IV diagnoses in the world mental health (WMH) version of the world health organization (WHO) Composite International Diagnostic Interview (WMH-CIDI). *International Journal of Methods in Psychiatric Research*, 13(2), 122-139.
- Kessler, R. C., Angermeyer, M., Anthony, J. C., de Graaf, R., Demyttenaere, K., Gasquet, I., de Girolamo, G., Gluzman, S., Gureje, O., Haro, J. M., Kawakami, N., Karam, A., Levinson, D., Mora, M. E. M., Browne, M. A. O., Posada-Villa, J., Stein, D. J., Tsang, C. H. A., Aguilar-Gaxiola, S., Alonso, J., Lee, S., Heeringa, S., Pennell, B. E., Berglund, P., Gruber, M. J., Petkhova, M., Chatterji, S. y Üstün, T. B. (2007). Lifetime prevalence and age-of-onset distributions of mental disorders in the World Health Organization's World Mental Health Survey Initiative. *World Psychiatry*, 6(3), 168-176.
- Kessler, R. C., Berglund, P., Demler, O. y Jin, R. (2003). The epidemiology of major depressive disorder: Results from the national comorbidity survey replication (NCS-R). *JAMA: Journal of American Medicine Association*, 289(23), 3095-3105.
- Kessler, R. C., Birnbaum, H. G., Shahly, V., Bromet, E., Hwang, I., McLaughlin, K. A., Sampson, N., Andrade, L. H., de Girolamo, G., Demyttenaere, K., Haro, J. M., Karam, A. N., Kostyuchenko, S., Kovess, V., Lara, C., Levinson, D., Matschinger, H., Nakane, Y., Browne, M.O., Ormel, J., Posada-Villa, J., Sagar, R. y Stein, D. J. (2010). Age differences in the prevalence and co-morbidity of DSM-IV major depressive episodes: results from the WHO World Mental Health Survey Initiative. *Depression and Anxiety*, 27(4), 351-364.

- Kessler, R. C., Chiu, W. T., Demler, O. y Walters, E. E. (2005). Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, 62(6), 617-627
- Kessler, R. C., McGonagle, K. A., Zhao, S., Nelson, C. B., Hughes, M., Eshleman, S., Wittchen, H. U. y Kendler, K. S. (1994). Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States. *Archives General of Psychiatry*, 51(1), 8-19.
- Kessler, R. C., Merikangas, K. R., Berglund, P., y Eaton, W. W. (2003). Mild disorders should not be eliminated from the DSM-V. *Archives of General Psychiatry*, 60(11), 1117-1122.
- Kessler, R. C., y Üstün, T. B. (2004). The World Mental Health (WMH) Survey Initiative version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI). *International Journal of Methods in Psychiatric Research*, 13(2), 93-121.
- Kirk, S. A. y Kutchins, H. (2002). Reliability and the remarkable achievement. En S. A. Kirk y H. Kutchins (Eds.), *The selling of DSM: the rhetoric of science in psychiatry* (pp. 133-166). New York: Transaction Publishers.
- Klerman, G. L. y Weissman, M. M. (1996). Increasing rates of depression. *JAMA: Journal of the American Medical Association*, 261(15), 2229-35.
- Knapp, M. K (2003). Hidden costs of mental illness. *British Journal of Psychiatry*, 183, 477-478.
- Kobak, K. A., Dottl, S. L., y Greist, J. H. (1997). A computer-administered telephone interview to identify mental disorders. *JAMA: Journal of American Medicine Association*, 278(11), 905-910.
- Kraemer, H., Kupfer, D., Clarke, D., Narrow, W. y Regier, D. (2012). DSM-5: How reliable is reliable enough? *American Journal of Psychiatry*, 169, 13-15.
- Kranzler, H. R., Kadden, R. M., Babor, T. F., Tennen, H., y Rounsaville, B. J. (1996). Validity of the SCID in substance abuse patients. *Addiction*, 91(6), 859-868.
- Kranzler, R., Kadden, R., Burleson, J., Babor, T. F., Apter, A., y Rounsaville, B. J. (1995). Validity of psychiatric diagnoses in patients with substance use



- disorders: Is the interview more important than the interviewer? *Comprehensive Psychiatry*, 36(4), 278-288.
- Kroenke, K., Spitzer, R. L., y Williams, J. B. (2001). The PHQ-9. *Journal of General Internal Medicine*, 16(9), 606-613.
- Kroenke K, Spitzer R. L. y Williams J. B. (2003). The Patient Health Questionnaire- 2: validity of a two-item depression screener. *Medical Care*, 41(11), 1284-1292.
- Krueger, R. F., y Markon, K. E. (2006). Reinterpreting comorbidity: a model-based approach to understanding and classifying psychopathology. *Annual Review of Clinical Psychology*, 2, 111-133.
- Krueger, R. F., Watson, D., y Barlow, D. H. (2005). Introduction to the special section: toward a dimensionally based taxonomy of psychopathology. *Journal of Abnormal Psychology*, 114(4), 491-493.
- Kurdyak, P. A. y Gnam, W. H. (2005). Small signal, big noise: performance of the CIDI depression module. *Canadian Journal of Psychiatry*, 50(13), 851-856.
- Labrador, F. J., Echeburúa, E., y Becoña, E. (2000). *Guía para la elección de tratamientos psicológicos efectivos. Hacia una nueva psicología clínica*. Madrid: Dykinson.
- Lecrubier, Y., Sheehan, D. V., Weiller, E., Amorim, P., Bonora, I., Sheehan, K. H., Janavs, J. y Dunbar, G.C. (1997). The Mini International Neuropsychiatric Interview (MINI). A short diagnostic structured interview: Reliability and validity according to the CIDI. *European Psychiatry*, 12(5), 224-231.
- Lenderink, A. F., Zoer, I., van der Molen, H. F., Spreeuwers, D., Frings-Dresen, M. H., y van Dijk, F. J. (2012). Review on the validity of self-report to assess work-related diseases. *International Archives of Occupational and Environmental Health*, 85(3), 229-251.
- Lépine, J. P., Gastpar, M., Mendlewicz, J. y Tylee, A., (1997). Depression in the community: The first pan-European study DEPRES (Depression Research in European Society). *International Clinical Psychopharmacology*, 12(1), 19-29.
- Levinson, D. F. (2009). Genetics of major depression. En I. H. Gotlib y C. L. Hammen (Eds.), *Handbook of depression, Second Edition* (pp. 165-186). New York: Guilford Press.

- Lewis, G., Pelosi, A. J., Araya, R. y Dunn, G. (1992). Measuring psychiatric disorder in the community; a standardised assessment for use by lay interviewers. *Psychological Medicine*, 22(2), 465-486.
- Lewinsohn, P. M., Petit, J. W., Joiner, T. E., y Seeley, J. R. (2003). The symptomatic expression of major depressive disorder in adolescents and young adults. *Journal of Abnormal Psychology*, 112(2), 244-252.
- Lipsman, N. y Lozano, A. M. (2011). The most cited works in major depression: The 'Citation classics'. *Journal of Affective Disorders*, 134(1-3), 39-44.
- Lishman, W.A. (1998). *Organic Psychiatry. The Psychological Consequences or Cerebral Disorder*. 3ª Edición, (161-217). Oxford: Black Well Science.
- Lobbestael, J., Leurgans, M. y Arntz, A. (2011). Inter-rater reliability of the Structured Clinical Interview for DSM-IV Axis I Disorders (SCID I) and Axis II Disorders (SCID II). *Clinical Psychology and Psychotherapy*, 18(1), 75-79.
- Lobo, A., Chamorro, L., Luque, A., Dal-Re, R., Badia, X. y Baro, E. (2002). Validación de las versiones en español de la Montgomery-Asberg Depression Rating Scale y la Hamilton Anxiety Rating Scale para la evaluación de la depresión y de la ansiedad. *Medicina Clínica*, 118(13), 493-499.
- López, A. L., Mathers, C. D., Ezzati, M., Jamison., D. T. y Murray, J. L. (2006). Global and regional burden of disease and risk factors, 2001: systematic analysis of population health data. *Lancet*, 367 (9524), 1747-1757.
- Lothgren, M. C. (2004). Economic evidence in affective disorders; a review. *European Journal of Health Economics*, 5(Suppl.1), 12-20.
- MacMillan, H. L., Patterson, C. J. S. y Wathen, C. N. (2005). Canadian Task Force on Preventive Health Care. Screening for depression in primary care: recommendation statement from the Canadian Task Force on Preventive Health Care. *Canadian Medical Association Journal*, 172, 33-35.
- Maj, M., Gaebel, W., López-Ibor, J. J., y Sartorius, N. (2002). Preface. En M. Maj, W. Gaebel, J. J. López-Ibor y N. Sartorius (Eds.), *Psychiatric diagnosis and classification* (pp. ix-xi). Chichester, UK: John Wiley & Sons.

- Martínez-Azumendi, O., Fernández-Gómez, C., y Beitia-Fernández, M. (2001). Variabilidad factorial del SCL-90-R en una muestra psiquiátrica ambulatoria. *Actas Españolas De Psiquiatría*, 29(2), 95-102.
- Meehl, P. E. (2010). Diagnostic Taxa as Open Concepts: metatheoretical and statistical questions about reliability and construct validity in the Grand Strategy of Nosological Revision. En T. Millon, R. F. Frueger, E. Simonsen (Eds.), *Contemporary directions in psychopathology: scientific foundations of the DSM-V and ICD-10* (pp. 174-186). New York: Guilford Press.
- Millon, T. (1974). *Psicopatología y personalidad*, 2ª ed. México, D. F.: Interamericana.
- Millon, T. (1976). *Psicopatología moderna: enfoque biosocial de los aprendizajes erróneos y de los disfuncionalismos*. Barcelona: Salvat.
- Millon, T. (1983). *Millon Clinical Multiaxial Inventory manual* (3.ª ed). Minneapolis. MN: National Computer Systems.
- Millon, T. (1987). *Millon Clinical Multiaxial Inventory-II manual* (2ª ed.). Minneapolis. MN: National Computer Systems.
- Millon, T. (1994). *The Millon Clinical Multiaxial Inventory-III manual*. Minneapolis, MN: National Computer Systems.
- Millon, T. (1998). *MCMI-II. Inventario Clínico Multiaxial de Millon-II. Manual* (Adaptación española: Ávila-Espada, A.). Madrid: TEA Ediciones.
- Millon, T. (1999). *MCMI-II. Inventario Clínico Multiaxial de Millon II. Manual* (2.ª ed.) (Adaptación española: Ávila-Espada, A.). Madrid: TEA Ediciones.
- Millon, T. (2002). *MCMI-II. Inventario Clínico Multiaxial de Millon II. Manual* (3.ª ed.) (Adaptación española: Ávila-Espada, A.). Madrid: TEA Ediciones.
- Millon, T. (2008). A brief history of psychopathology. En P. H. Blaney y T. Millon (Eds.), *Oxford textbook of psychopathology* (pp. 3-34). New York: Oxford University Press.
- Millon, T., y Bloom, C. (Eds.) (2008). *The Millon inventories: a practitioner's guide to personalized clinical assessment*. New York: Guilford Press.
- Millon, T. y Davis, R. D. (1997). The MCMI-III: present and future directions. *Journal of Personality Assessment*, 68(1), 69-85.

- Millon, T. y Davis, R. D. (1998). *Trastornos de la personalidad. Más allá del DSM-IV*. Barcelona: Masson.
- Millon, T., Davis, R. D., y Millon, C. (1997). *The Millon Clinical Multiaxial Inventory-III manual* (2ª ed.). Minneapolis, MN: National Computer Systems.
- Millon, T., Davis, R. D., y Millon, C. (2007). *MCMI-III. Inventario Clínico Multiaxial de Millon-III. Manual* (Adaptación española: Cardenal, V., y Sánchez, M. P.). Madrid: TEA Ediciones.
- Millon, T., Millon, C., Davis, R., y Grossman, S. (2009). *MCMI-III. Millon Clinical Multiaxial Inventory-III. Manual* (4ª ed.). San Antonio, Texas: Pearson Clinical Assessment.
- Millon, T., y Simonsen, E. (2010). A précis of psychopathological history. En T. Millon, R. F. Frueger y E. Simonsen (Eds.), *Contemporary directions in psychopathology: scientific foundations of the DSM-V and ICD-10* (pp. 3-52). New York: Guilford Press.
- Mitchell, A. J., Rao, S. y Vaze, A. (2011). Can general practitioners identify people with distress and mild depression? A meta-analysis of clinical accuracy. *Journal of Affective Disorders*, 130(1-2), 26-36.
- Montgomery, S. A. y Asberg, M. (1979). A new depression scale designed to be sensitive to change. *The British Journal of Psychiatry*, 134, 382-389.
- Mordal, J., Gundersen, O. y Bramness, J. G. (2010). Norwegian version of the Mini-International Neuropsychiatric Interview: Feasibility, acceptability and test-retest reliability in an acute psychiatric. *European Psychiatry*, 25(3), 172-177.
- Mullins-Sweatt, S. N. y Widiger, T. A. (2009). Clinical Utility and DSM-V. *Psychological Assessment*, 21(3), 302-312.
- Muñoz, M. (2003). Evaluación para el diagnóstico. En *Manual Práctico de la Evaluación Psicológica Clínica* (pp. 59-94). Madrid: Editorial Síntesis.
- Muñiz, J., y Fernández-Hermida, J. R. (2010). La opinión de los psicólogos españoles sobre el uso de los tests. *Papeles del Psicólogo*, 31(1), 108-121.
- Murray, J. L. y López, A. D., (1997). Global mortality, disability, and the contribution of risk factors: Global Burden of Disease Study. *The Lancet*, 349(9063), 1436-1442.

- Narrow W. E., Rae D. S., Robins L. N. y Regier D. A. (2002). Revised prevalence estimates of mental disorders in the United States: using a clinical significance criterion to reconcile 2 surveys' estimates. *Archives of General Psychiatry*, 59(2), 115-23.
- Nezu, A. M., Nezu, C. M., Friedman, J. y Lee, M. (2009). Assessment of Depression. En Gotlib, I. H. y Hammen C. L. (Eds.), *Handbook of depression, Second Edition* (pp. 44-68). New York: The Guilford Press.
- NICE, National Institute of Clinical Excellence (2010). *Depression: The Guideline on the Treatment and Management of Depression in Adults National Collaborating Centre for Mental Health (NCCMH)*. Royal College of Psychiatrists y British Psychological Society.
- Nienhuis, F. J., van, d. W., Rijnders, C. A. T., de Jonge, P., y Wiersma, D. (2010). Validity of a short clinical interview for psychiatric diagnosis: The mini-SCAN. *British Journal of Psychiatry*, 196(1), 64-68.
- Nolen-Hoeksema, S. y Hilt, L. M. (2009). Gender Differens in depression. En I. H. Gotlib, y C. L. Hammen (Eds.), *Handbook of depression, Second Edition* (pp 386-404). New York: The Guilford Press.
- Norcross, J. C., y Wampold, B. E. (2011). *Evidence-based therapy relationships: Research conclusions and clinical practices. Psychotherapy*, 48(1), 98-102.
- Oliva-Moreno, J., López-Bastida, J., Montejo-González, A. L., Osuna-Guerrero, R. y Duque-González, B. (2009). The socioeconomic costs of mental illness in Spain. *European Journal of Health Economic*, 10(4), 361-369.
- Organización Mundial de la Salud (1993) *CIE -10. Trastornos Mentales y del Comportamiento*. Madrid: Meditor.
- Organización Mundial de la Salud (1997). Entrevista Diagnóstica Internacional Compuesta (CIDI) Versión 2.1.
- Organización Mundial de la Salud (2000). *Preventing Suicide. A Resource for General Physician*. Genova. Recuperado el 21 de febrero de 2012. [http://www.who.int/mental\\_health/media/en/56.pdf](http://www.who.int/mental_health/media/en/56.pdf).
- Organización Mundial de la Salud (2004). The global burden of disease: 2004 update. Recuperado el 21 de febrero de 2012.

[http://www.who.int/healthinfo/global\\_burden\\_disease/2004\\_report\\_update/en/index.html](http://www.who.int/healthinfo/global_burden_disease/2004_report_update/en/index.html).

- Orive, M., Padierna, J. A., Quintana, J. M., Las-Hayas, C., Vrotsou, K., y Aguirre, U. (2010). Detecting depression in medically ill patients: comparative accuracy of four screening questionnaires and physicians' diagnoses in Spanish population. *Journal of Psychosomatic Research*, 69(4), 399-406.
- Pablo Alcázar, F. de, Albacete Belmonte, A., Meroño Méndez, A., Artiz Martínez, M, Berbell Torrecilla, F., García Peñalver, A., Giménez Giménez, L. , Martínez de Salazar Arboleas, A., Mateu Sabater, A., Plaza Tobarra, R., Sepulcre Canto, R. y Sánchez Hurtado, N. (1992). Validez de la versión española del Diagnostic Interview Schedule. Tercera Edición Revisada (DIS-III-R). *Actas Luso-Españolas de Neurología, Psiquiatría y Ciencias Afines*, 20(6), 257-262
- Patten, S. B. y Schopflocher, D. (2009). Longitudinal epidemiology of major depression as assessed by the Brief Patient Health Questionnaire (PHQ-9). *Comprehensive Psychiatry*, 50(1), 26-33.
- Paykel, E. S. (2006). Editorials: Depression: Major problem for public health. *Epidemiologia e Psichiatria Sociale*, 15(1), 4-10.
- Pérez, M., Fernández, J. R., Fernández, C. y Amigo, I. (2005). El fin de la inocencia en los tratamiento psicológicos. Análisis de la situación actual. En M., Pérez, J. R., Fernández, C. Fernández, y I. Amigo (Eds.), *Guía de los tratamientos psicológicos eficaces I* (pp. 17-34). Madrid: Editorial Pirámide.
- Phillips C. J., Henderson, A. S. (1991). The prevalence of depression among Australian nursing home residents: results using draft ICD-10 and DSM-III-R criteria. *Psychological Medicine*, 21(03), 739-748.
- Piersma, H. L. (1991). The MCMI-II depression scales: Do they assist in the differential prediction of depressive disorders? *Journal of Personality Assessment*, 56(3), 478-486.
- Pignone, M. P., Gaynes, B. N., Rushton, J. L., Burchell, C. M., Orleans, C. T., Mulrow, C. D. y Lohr, K. N. (2002). Screening for depression in adults: a summary of the evidence for the US Preventive Services Task Force. *Annals of Internal Medicine*, 136(10), 765-76.

- Pincus, H. A., McQueen, L. E. y Elinson, L., (2005). Trastornos mentales subliminales. En K. A., Phillips, M. B., First, H. A., Pincus (Eds.), *Avances en el DSM: dilemas en el Diagnóstico Psiquiátrico* (pp. 127-142). Barcelona: Editorial Masson.
- Prieto, G., y Muñoz, J. (2000). Un modelo para evaluar la calidad de los tests utilizados en España. *Papeles del Psicólogo*, 77, 65-71.
- Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, 1(3), 385-401.
- Ramos-Brieva, J. y Cordero Villafáfila, A. (1986). Validación de la versión castellana de la escala de Hamilton para la depresión. *Actas Luso-Espanolas De Neurología, Psiquiatria y Ciencias Afines*, 14(4), 324-334.
- Ramos-Brieva, J., Cordero Villafáfila, A. y Yañez Sáez, R. (1994). Nuevos datos sobre la validez y fiabilidad de la versión castellana de la Escala de Hamilton para la Depresión. *Anales De Psiquiatría*, 10(4), 146-151.
- Regier, D. A., Narrow, W. E., Clarke D. E., Kraemer, H. C., Kuramoto, S. J., Kuhl, E. A. y Kupfer, D. J. (2013) DSM-5 Field Trials in the United States and Canada, Part II: Test-Retest Reliability of Selected Categorical Diagnoses. *American Journal of Psychiatry*, 170(1), 59-70.
- Regier, D. A., Narrow, W. E., Kuhl, E. A., y Kupfer, D. J. (2009). The conceptual development of DSM-V. *The American Journal of Psychiatry*, 166(6), 645-650.
- Retzlaff, P. D. (1996). MCMI-III diagnostic validity: Bad test or bad validity study. *Journal of Personality Assessment*, 66(2), 431-437.
- Retzlaff, P. D., Sheehan, E. P. y Lorr, M. (1990). MCMI-II scoring: Weighted and unweighted algorithms. *Journal of Personality Assessment*, 55(1-2), 219-223.
- Rijnders, C. A. T., van den Berg, J. F., Hodiament, P. P. G., Nienhuis, F. J., Furer, J. W., Mulder, J. y Giel, R. (2000). Psychometric properties of the Schedules for Clinical Assessment in Neuropsychiatry (SCAN-2.1). *Social Psychiatry and Psychiatric Epidemiology*, 35(8), 348-352.
- Robins, L. N., Cottler, L. B., Bucholz, K., Compton, W. M., North, C. S. y Rourke, K. M. (2001). *Diagnostic Interview Schedule-IV*. Minneapolis: National Computer Systems.

- Robins, L. N., Cottler, L. B., Bucholz, K. K., Compton, W. M., North, C. S. y Rourke, K. M. (1995). Diagnostic Interview Schedule for DSM-IV (DIS-IV). St. Louis: Washington University.
- Robins, L. N., Helzer, J. E., Croughan, J. L., y Ratcliff, K. S. (1981). National Institute of Mental Health Diagnostic Interview Schedule: Its history, characteristics, and validity. *Archives of General Psychiatry*, 38(4), 381-389.
- Roca, M., Gili, M., Ferrer, V., Bernardo, M., Montaña, J. J., Salvà, J. J., Flores, I. y Leal, S. (1999). Mental disorders on the island of Formentera: Prevalence in general population using the schedules for clinical assessment in neuropsychiatry (SCAN). *Social Psychiatry and Psychiatric Epidemiology*, 34(8), 410-415.
- Roca, M., Gili, M., Ferrer, V., y Bernardo, M. (2001). Mental disorders and medical conditions. A community study in a small island in Spain. *Journal of Psychosomatic Research*, 50(1), 39-44.
- Rogers, R. (2003). Standardizing DSM-IV diagnoses: The clinical applications of structured interviews. *Journal of Personality Assessment*, 81(3), 220-225.
- Rogers, R. (2001). Schedule for Affective Disorders and Schizophrenia. En *Handbook of diagnostic and structured interviewing*, (pp. 84-102). New York: Guilford Press.
- Rogers, R., Salekin, R. T., y Sewell, K. W. (1999). Validation of the Millon Clinical Multiaxial Inventory for axis II disorders: Does it meet the Daubert standard? *Law and Human Behavior*, 23(4), 425-443.
- Rossi, A., Alberio, R., Porta, A., Sandri, M., Tansella, M., y Amaddeo, F. (2004). The reliability of the Mini-International Neuropsychiatric Interview- Italian Version. *Journal of Clinical Psychopharmacology*, 24(5), 561-563.
- Rossi, G., y Sloore, H. (2005). Diagnostic efficiency of the Dutch-language version of the Millon Clinical Multiaxial Inventory-III. En I. V. Halvorsen y S. N. Olsen (Eds.), *New research on personality disorders* (pp. 193-211). New York: Nova Science Publishers.
- Saladrigas, M. V., y Sacristán del Castillo, J. A. (2004). Fichas de MedTrad (n.º 13): eficacia, efectividad y eficiencia en la investigación de fármacos. *Panace@*, 16, 97-99.
- Santor, D. A., Gregus, M. y Welch, A. (2006). Eight decades of measurement in



- depression. *Measurement: Interdisciplinary Research & Perspective*, 4(3), 135-155.
- Santor, D. A., y Coyne, J. C. (1997). Shortening the CES-D to improve its ability to detect cases of depression. *Psychological Assessment*, 9(3), 233-243.
- Sanz, J. (2013). *Propiedades psicométricas de un instrumento para la heteroevaluación de la depresión: el Listado para el Episodio Depresivo (LED)*. Manuscrito sin publicar. Facultad de Psicología. Universidad Complutense de Madrid.
- Sanz, J., y García-Vera, M. P. (2009). The Beck Depression Inventory-second edition (BDI-II): factor congruence and generalizability of its indexes of internal consistency. En E. Řehulka (Ed.), *School and health 21. General issues in health education* (pp. 331-342). Brno: MSD.
- Sanz, J., y García-Vera, M. P. (2013). Rendimiento diagnóstico y estructura factorial del Inventario de Depresión de Beck-II (BDI-II). *Anales de Psicología*, 29, 66-75.
- Sanz, J., García-Vera, M. P., Espinosa, R., Fortún, M., y Vázquez, C. (2005). Adaptación española del Inventario para la Depresión de Beck-II (BDI-II): 3. Propiedades psicométricas en pacientes con trastornos psicológicos. *Clínica y Salud*, 16, 121-142.
- Sanz, J., Izquierdo, A., y García-Vera, M. P. (en prensa). Una revisión desde la perspectiva de la validez de contenido de los cuestionarios autoaplicados utilizados en España para evaluar la depresión clínica en adultos. *Psicopatología Clínica, Legal y Forense*.
- Sanz, J., Navarro, M. E., y Vázquez, C. (2003). Adaptación española del Inventario para la Depresión de Beck-II (BDI-II): 1. Propiedades psicométricas en estudiantes universitarios. *Análisis y Modificación de Conducta*, 29, 239-288.
- Sanz, J., Perdigón, A. L., y Vázquez, C. (2003). Adaptación española del inventario para la depresión de beck-II (BDI-II): 2. Propiedades psicométricas en población general. *Clínica y Salud*, 14(3), 249-280.
- Saulsman, L. M. (2011). Depression, anxiety, and the MCMI-III: Construct validity and diagnostic efficiency. *Journal of Personality Assessment*, 93(1), 76-83.

- Schmitz, N., Kruse, J., Heckrath, C., Alberti, L., y Tress, W. (1999). Diagnosing mental disorders in primary care: The General Health Questionnaire (GHQ) and the Symptom Check List (SCL-90-R) as screening instruments. *Social Psychiatry and Psychiatric Epidemiology*, 34(7), 360-366.
- Scott P. B. (2008). The evolving understanding of major depression epidemiology: Implications for practice and policy. *Canadian Journal of Psychiatry*, 53(10), 689-695.
- Segal, D. L., Coolidge, F. L., O'Riley, A. y Heinz, B. A. (2006). Structured and semistructured interviews. En M. Hersen, (Ed.), *Clinician's handbook of adult behavioral assessment* (pp. 121-144). Boston: Elsevier Academic Press.
- Segal, D. L., June, A. y Marty, M. A. (2010). Basic issues in interviewing and the interview process. En D. L., Segal, M. Hersen (Eds.), *Diagnostic interviewing*, (pp. 1-22). New York: Springer Publishing.
- Sheehan, D., Lecrubier, Y., Harnett-Sheehan, K., Janavs, J., Weiller, E., Kesinker, A., Schinka, J., Knapp, E., Sheehan, M. F., y Dunbar, G. (1997). The validity of the Mini International Neuropsychiatric Interview (MINI) according to the SCID-P and its reliability. *European Psychiatry*, 12, 232-241.
- Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., Hergueta, T., Baker, R., y Dunbar, G. C. (1998). The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *Journal of Clinical Psychiatry*, 59(Supl. 20), 22-33.
- Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Janavs, J., Weiller, E., Keskiner, A., Schinka, J., Knapp, E., Sheehan, M. F. y Dunbar G. C. (1997). The validity of the Mini International Neuropsychiatric Interview (MINI) according to the SCID-P and its reliability. *European Psychiatry*, 12(5), 232-241.
- Sheeran, T., y Zimmerman, M. (2002). Case identification of depression with self-report questionnaires. *Psychiatry Research*, 109(1), 51-59.
- Shorter, E. (2007). The doctrine of the two depressions in historical perspective. *Acta Psychiatrica Scandinavica*, 115(Suppl. 433), 5-13.

- Slade, T. y Sunderland, M. (2010). Quantifying point prevalence of major depressive episode using lifetime structured diagnostic interviews. *Journal of Affective Disorders*, 121, 39-44.
- Smith, G. T. y Combs, J. (2010). Issues of construct validity in psychological diagnoses. En T., Millon R. F. Krueger y E. Simonsen (Eds.). *Contemporary directions in psychopathology: scientific foundations of the DSM-V and ICD-11*, pp. 205-222). New York, US: Guilford Press.
- Sobocki, P., Jönsson, B., Angst, J. y Rehnberg, C. (2006). Cost of depression in Europe. *Journal of Mental Health Policy and Economics*, 9(2), 87-98.
- Sobocki, P. y Wittchen, H.-U. (2005). Cost of affective disorders in Europe. *European Journal of Neurology*, 12(Suppl. 1), 34-38.
- Soler, J., Pérez Sola, V., Puigdemot, D, Pérez-Blanco, J., y Figueres, M. (1997). Estudio de validación del *Center for Epidemiologic Studies–Depression* (CESD) en una población española de pacientes con trastornos afectivos. *Actas Luso-Españolas de Neurología, Psiquiatría y Ciencias afines*, 25, 243-249.
- Sox, H., Stern, S., Owens, D., y Abrams, H. L. (1989). *Assessment of diagnostic technology in health care. Rationale, methods, problems, and directions*. Washington: National Academic Press.
- Spitzer, R. L. (1983). Psychiatric diagnosis: Are clinicians still necessary? *Comprehensive Psychiatry*, 24, 399-411.
- Spitzer, R. L. (2001). Values and assumptions in the development of DSM-III and DSM-III-R: an insider's perspective and a belated response to Sadler, Huls, and Agich's "On values in recent American psychiatric classification". *Journal of Nervous and Mental Disease*, 189, 351-359.
- Spitzer, R. L., Endicott, J., y Robins, E. (1978). Research diagnostic criteria: rationale and reliability. *Archives of General Psychiatry*, 35(6), 773-782.
- Spitzer, R. L., Kroenke, K., y Williams, J. B. (1999). Validation and utility of a self-report version of PRIME-MD: The PHQ primary care study. *JAMA: Journal of American Medicine Association*, 282, 1737-1744.
- Spitzer, R. L., y Wakefield, J. C. (1999). DSM-IV diagnostic criterion for clinical significance: Does it help solve the false positives problem? *The American*

*Journal of Psychiatry*, 156(12), 1856-1864.

- Spitzer, R. L., Williams, J. B. W., Kroenke, K., Linzer, M., de Gruy III, F. V., Hahn, S. R., Brody, D., y Johnson, J. G. (1994). Utility of a new procedure for diagnosing mental disorders in primary care: The PRIME-MD 1000 study. *JAMA: Journal of the American Medical Association*, 272, 1749-1756.
- Steer, R. A., Ball, R., Ranieri, W. F., y Beck, A. T. (1997). Further evidence for the construct validity of the Beck Depression Inventory-II with psychiatric outpatients. *Psychological Reports*, 80(2), 443-446.
- Steer, R. A., Cavalieri, T. A., Leonard, D. M., y Beck, A. T. (1999). Use of the Beck Depression Inventory for Primary Care to screen for major depression disorders. *General Hospital Psychiatry*, 21(2), 106-111.
- Stengel, D., y Porzsolt, F. (2006). Efficacy, effectiveness, and efficiency of diagnostic technology. En F. Porzsolt y R. M. Kaplan (Eds.), *Optimizing health: improving the value of healthcare delivery* (pp. 217-231). Nueva York: Springer Science+Business Media.
- Streiner, D. L., y Cairney, J. (2007). What's under the ROC? An introduction to Receiver Operating Characteristics curves. *Canadian Journal of Psychiatry*, 52, 121-128.
- Streiner, D. L., Goldberg, J. O. y Miller, H. R. (1993). MCMI-II item weights: their lack of effectiveness. *Journal of Personality Assessment*, 60(3), 471-476.
- Summerfeldt, L. J., Klossterman, P. H. y Antony, M. (2011). Structured and semistructured diagnostic interviews. En M. Antony y D. H. Barlow (Eds.), *Handbook of assessment and treatment planning for psychological disorders* (pp. 95-137). New York: Guilford Press.
- Swets, J. (1988). Measuring the accuracy of diagnostic systems. *Science*, 240, 1285-1293.
- Symposium. Melancholia: Beyond DSM, Beyond Neurotransmitters, May 2-4, 2006, Copenhagen Marriott. (2007). Transcript of discussions. *Acta Psychiatrica Scandinavica*, 115(Suppl. 433), 136-183.
- Thompson, C., Ostler, K., Peveler, R. C., Baker, N. y Kinmonth, A. L. (2001). Dimensional perspective on the recognition of depressive symptoms in primary

- care. The Hampshire Depression Project 3. *British Journal of Psychiatry*, 179, 317-323.
- Titov, N., Dear, B. F., McMillan, D., Anderson, T., Zou, J. y Sunderland, M. (2011). Psychometric comparison of the PHQ-9 and BDI-II for measuring response during treatment of depression. *Cognitive Behaviour Therapy*, 40(2), 126-136.
- Torrens, M., Serrano, D., Astals, M., Perez-Dominguez, G., y Martin-Santos, R. (2004). Diagnosing comorbid psychiatric disorders in substance abusers: validity of the Spanish versions of the Psychiatric Research Interview for Substance and Mental Disorders and the Structured Clinical Interview for DSM-IV. *The American Journal of Psychiatry*, 161(7), 1231-1237.
- Trajković, G., Starčević, V., Latas, M., Leštarević, M., Ille, T., Bukumirić, Z. y Marinković, J. (2011). Reliability of the Hamilton Rating Scale for depression: A meta-analysis over a period of 49 years. *Psychiatry Research*, 189(1), 1-9.
- U.S. Preventive Services Task Force (2009). *Screening for Depression in Adults: Recommendation Statement*. AHRQ Publication No. 10-05143-EF-2. Recuperado el 21 de febrero de 2012. <http://www.uspreventiveservicestaskforce.org/uspstf09/adultdepression/addeprss.htm>
- Üstün T. B., Ayuso-Mateos J. L., Chatterji S., Mathers C. y Murray C. J. L. (2004). Global burden of depressive disorders in the year 2000. *British Journal of Psychiatry*, 184, 386-392.
- Üstün, T. B., Chatterji, S. y Andrews, G. (2002). International classifications and the diagnosis of mental disorders: strengths, limitations and future perspectives. En M. Maj, W. Gaebel, J. J. López-Ibor y N. Sartorius (Eds.), *Psychiatric diagnosis and classification* (pp. 25-46). Chichester, UK: John Wiley & Sons.
- Valero, L., Ruiz M. A. (2003). Evaluación de un servicio de salud mental: análisis de la demanda y datos epidemiológicos. *Revista de Psiquiatría, Psicología Médica y Psicosomática*, 24(1), 15-22.
- Vázquez-Barquero, J. L., Gaité, L., Artal, J., Herrera S., Díez-Manrique, J. F, Cuesta M.J. e Higuera, A. (1994). Desarrollo y verificación de la versión española de la entrevista psiquiátrica "sistema SCAN" (Cuestionarios para la evaluación clínica en neuropsiquiatría). *Actas Luso Españolas Psiquiatría, Neurología y Ciencias Afines*, 22, 109-120.

- Vázquez, C. (1990). Fundamentos teóricos y metodológicos de la clasificación en psicopatología. En F. Fuentenebro y C. Vázquez (Eds.), *Psicología médica, psicopatología y psiquiatría. Volumen II: psiquiatría y psicología clínicas* (pp. 655-678). Madrid: McGraw Hill.
- Vázquez, F. L. y Blanco, V. (2008). Prevalence of DSM-IV major depression among Spanish university students. *Journal of American College Health*, 57(2), 165-171.
- Vázquez, F. L., Blanco, V., y López, M. (2007). An adaptation of the Center for Epidemiologic Studies Depression Scale for use in non-psychiatric Spanish populations. *Psychiatry Research*, 149(1-3), 247-252.
- Vázquez, C., Hervás, G., Rahona, J. J. y Gómez, D. (2009). Psychological well-being and health: contributions from positive psychology. *Annuary of Clinical and Health Psychology*, 5, 15-28.
- Vázquez, C., Sánchez, A. y Romero, N. (2011). Clasificación y diagnóstico de los trastornos psicológicos: el DSM-IV-TR/DSM-V y la CIE-10. En V. E. Caballo, I. C. Salazar y J. A. Carroble (Eds.), *Manual de psicopatología y trastornos psicológicos* (pp. 92-105). Madrid: Pirámide.
- Vázquez, C., y Sanz, J. (1991). Trastornos depresivos (I): datos clínicos y modelos teóricos. En A. Belloch y E. Ibáñez (Eds.), *Manual de psicopatología* (vol. II, pp. 717-784). Valencia: Promolibro.
- Vázquez, C., y Sanz, J. (2008). Trastornos del estado de ánimo: Aspectos clínicos. En A. Belloch, B. Sandín, y F. Ramos (Eds.), *Manual de psicopatología. Vol. 2, 2ª ed. revisada* (pp. 233-269). Madrid: McGraw Hill.
- Vizcarro, C., (2000). Clasificación y criterios diagnósticos de la conducta anormal. En R. Fernández Ballesteros (Ed.), *Introducción a la evaluación psicológica II* (pp. 102-123). Madrid: Pirámide.
- Wakefield, J. C. y First, M. B. (2005). Clarificación de la distinción entre lo que es y lo que no es trastorno. Afrontamiento del problema del sobrediagnóstico (falsos positivos) en el DSM-V. En K. A. Phillips, M. B. First y H. A. Pincus (Eds.), *Avances en el DSM: dilemas sobre el diagnóstico psiquiátrico* (pp. 23-55). Barcelona: Masson.

- Wakefield, J. C., Schmitz, M. F. y Baer, J. C. (2010). Does the DSM-IV clinical significance criterion for major depression reduce false positives? Evidence from the National Comorbidity Survey Replication. *American Journal of Psychiatry*, 167, 298-304.
- Wakefield, J. C., Schmitz, M. F., First, M. B. y Horwitz, A. V. (2007). Extending the bereavement exclusion for major depression to other losses: evidence from the National Comorbidity Survey. *Archives of General Psychiatry*, 64, 433-440.
- Wang, P. S., Simon, G. y Kessler, R. C. (2003). The economic burden of depression and the cost-effectiveness of treatment. *International Journal of Methods in Psychiatry Research*, 12(1), 22-33.
- Wang, J. C., Williams, J., Lavorato, D., Schmitz, N., Dewa, C. y Patten, S. B. (2010). The incidence of major depression in Canada: The National Population Health Survey. *Journal of Affective Disorders*, 123(1-3), 158-163.
- Wetzler, S. (1990). The Millon Clinical Multiaxial Inventory (MCMI): a review. *Journal of Personality Assessment*, 55(3-4), 445-464.
- Wetzler, S., Hahn, R., Strauman, T. J. y Dubro, A. (1989). Diagnosis of mayor depression by self-report. *Journal of Personality Assessment*, 53(1), 22-30.
- Wetzler, S. y Marlowe, D. B. (1993). The diagnosis and assessment of depression, mania, and psychosis by self-report. *Journal of Personality Assessment*, 60, 1-31.
- Williams, J. B. W., Gibbon, M., First, M. B., Spitzer, R. L., Davis, M., Borus, J., Howes, M. J., Kane, J., Pope, H. G., Rounsaville, B. y Wittchen, H.-U. (1992). The Structured Clinical Interview for DSM-III-R (SCID) II. Multi-site test-retest reliability. *Archives of General Psychiatry*, 49(8), 630-636.
- Wing, J. K. (1998). Aims and structure of SCAN. En J.K., Wing, N., Sartorius, T.B. Üstün (Eds.), *Diagnosis and clinical measurement in psychiatry: a reference manual for SCAN* (pp. 25-35). London: Cambridge University Press.
- Wing, J. K., Cooper, J. E. y Sartorius, N. (1974). Measurement and classification of psychiatric symptoms. London: Cambridge University Press.

- Wittchen, H.-U. y Jacobi, F. (2005) Size and burden of mental disorders in Europe -a critical review and appraisal of 27 studies. *European Neuropsychopharmacology*, 15, 357-376.
- Wittchen, H.-U. y Pittrow, D. (2002). Prevalence, recognition and management of depression in primary care in Germany: the Depression 2000 study. *Human Psychopharmacology*, 17, 1-11.
- Yin, P. y Fan, X. (2000). Assessing the reliability of Beck Depression Inventory Scores: reliability generalization across studies. *Educational and Psychological Measurement*, 60 (2), 201-223.
- Zachar, P. (2008). Real kinds but no true taxonomy: an eassay in psychiatric systematic. En K. S. Von Kendler y J. Parnas, (Eds.), *Philosophical issues in psychiatry: explanation, phenomenology, and nosology*, (pp. 327-367) Baltimore: The Johns Hopkins University Press.
- Zimmerman, M., Posternak, M. A. y Chelminski, I. (2005). Is it time to replace the Hamilton Depression Rating Scale as the primary outcome measure in treatment studies of depression? *Journal of Clinical Psychopharmacology*, 25(2), 105-110.
- Zimmerman, M., Sheeran, T. y Young D. (2004). The Diagnostic Inventory for Depression: a self-report scale to diagnose DSM-IV major depressive disorder. *Journal of Clinical Psychology*, 60(1), 87-110.
- Zisook, S. y Kendler, K. S. (2007). Is bereavement-related depression different than non-bereavement-related depression. *Psychological Medicine*, 37(6), 779-794.
- Zung W. W. (1965). A self-rating depression scale. *Archives of General Psychiatry*, 12, 63-70.
- Zunzunegui Pastor, M. V., Delgado, M., Pérez Pérez, E., Yagüe Alonso, A. I., Illescas, M. L., y Galat León, V. (1998). Validación de la Escala CES-D para la medida de la sintomatología depresiva en una población de personas mayores española. *Revista Multidisciplinar de Gerontología*, 8, 156-161.